# NOVEL DNAs AND POLYPEPTIDES

## Reference to Related Application

This application claims the benefit of U.S. Provisional Application S.N. 60/107,821, filed November 10, 1998, which is incorporated herein by reference.

## BACKGROUND OF THE INVENTION

### Field of the Invention

The invention is directed to purified and isolated novel polypeptide molecules and fragments thereof, the nucleic acid molecules encoding such polypeptides, processes for production of recombinant forms of such polypeptide molecules, antibodies generated against such polypeptide molecules, fragmented peptides derived from these polypeptide molecules, and uses thereof. In particular, the invention is directed to the use of the nucleic acid and polypeptide molecules in the study of inflammatory bowel diseases (IBD)

### Description of Related Art

Damage to the intestinal epithelial barrier is a hallmark of inflammatory bowel diseases (IBD). Examples of inflammatory bowel diseases include ileitis, Crohn's disease (CD), which can affect the whole digestive tract from mouth to anus, and ulcerative colitis (UC), which affects only the large intestine. Studying the factors that influence the integrity of the epithelial barrier *in vivo* is a difficult task for a number of reasons that include the complexity of the tissue itself (there are numerous cell types in the gut including epithelial, stromal, endocrine, neuronal and hematopoietic) and the technical problems associated with tissue manipulation in animals or in isolated organs. As a result of these issues, a number of *in vitro* models of epithelial barrier function have been developed over the years. The best characterized of these models is the T84 intestinal epithelial barrier system (Dharmsathaphorn et al., *Am. J. Physiol.*, 246:G204-G208, 1984 and Madara et al., *J. Cell Biol.*, 101:2124-2133, 1985). T84 cells, while derived from a human colonic adenocarcinoma, have retained many of the properties associated with normal colonic crypt cells. T84 cells form polarized monolayers that exhibit high electrical resistance and vectorial fluid and chloride secretion reminiscent of colonic crypts *in vivo*. These properties are directly dependent on a complex of proteins referred to as tight junctions and T84 cells express many of the known members of this complex (Youakim et al., *Submitted for*

*publication*, 1998). These cells also respond to proinflammatory cytokines, such as interferon-gamma (INF-gamma), by decreasing barrier function (Youakim et al., *Submitted for publication*, 1998; Madara et al., *J. Clin. Invest.*, 83:724-727, 1989; and Adams et al., *J. Immunol.*, 150:2356-2363, 1993) and by up-regulating MHC Class II molecules and antigen presenting activity (Hershberg et al., *J. Clin. Invest.*, 102:792-803, 1998). This model system is used to examine how the epithelial barrier is regulated by various agents such as interferon-gamma and cells of the immune system (e.g., neutrophils). The T84 model assists in the elucidation of the mechanism of barrier breakdown and recovery in response to these agents and in the identification of proteins (and genes) that may prevent barrier breakdown or stimulate barrier recovery.

Given the important function of the epithelial gut barrier and despite the growing body of knowledge, there is a need in the art for the discovery, the identification, and the elucidation of the roles of new proteins involved in gut barrier function and IBD.

The identification of the primary structure, or sequence, of an unknown protein is the culmination of an arduous process of experimentation. In order to identify an unknown protein, the investigator can rely upon a comparison of the unknown protein to known peptides using a variety of techniques known to those skilled in the art. For instance, proteins are routinely analyzed using techniques such as electrophoresis, sedimentation, chromatography, sequencing and mass spectrometry.

In particular, comparison of an unknown protein to polypeptides of known molecular weight allows a determination of the apparent molecular weight of the unknown protein (T.D. Brock and M.T. Madigan, *Biology of Microorganisms* 76-77 (Prentice Hall, 6d ed. 1991)). Protein molecular weight standards are commercially available to assist in the estimation of molecular weights of unknown protein (New England Biolabs Inc. Catalog:130-131, 1995; J. L. Hartley, U.S. Patent No. 5,449,758). However, the molecular weight standards may not correspond closely enough in size to the unknown protein to allow an accurate estimation of apparent molecular weight. The difficulty in estimation of molecular weight is compounded in the case of proteins that are subjected to fragmentation by chemical or enzymatic means, modified by post-translational modification or processing, and/or associated with other proteins in non-covalent complexes.

In addition, the unique nature of the composition of a protein with regard to its specific amino acid constituents results in unique positioning of cleavage sites within the protein. Specific fragmentation of a protein by chemical or enzymatic cleavage results in a unique "peptide fingerprint" (D. W. Cleveland et al., *J. Biol. Chem.* 252:1102-1106, 1977; M. Brown et

al., *J. Gen. Virol.* 50:309-316, 1980). Consequently, cleavage at specific sites results in reproducible fragmentation of a given protein into peptides of precise molecular weights. Furthermore, these peptides possess unique charge characteristics that determine the isoelectric pH of the peptide. These unique characteristics can be exploited using a variety of electrophoretic and other techniques (T.D. Brock and M.T. Madigan, *Biology of Microorganisms* 76-77 (Prentice Hall, 6d ed. 1991)).

Fragmentation of proteins is further employed for amino acid composition analysis and protein sequencing (P. Matsudiara, *J. Biol. Chem.* 262:10035-10038, 1987; C. Eckerskorn et al., *Electrophoresis 1988*, 9:830-838, 1988), particularly the production of fragments from proteins with a "blocked" N-terminus. In addition, fragmented proteins can be used for immunization, for affinity selection (R. A. Brown, U.S. Patent No. 5,151,412), for determination of modification sites (*e.g.* phosphorylation), for generation of active biological compounds (T.D. Brock and M.T. Madigan, *Biology of Microorganisms* 300-301 (Prentice Hall, 6d ed. 1991)), and for differentiation of homologous proteins (M. Brown et al., *J. Gen. Virol.* 50:309-316, 1980).

In addition, when a peptide fingerprint of an unknown protein is obtained, it can be compared to a database of known proteins to assist in the identification of the unknown protein using mass spectrometry (W.J. Henzel et al., *Proc. Natl. Acad. Sci. USA* 90:5011-5015, 1993; D. Fenyo et al., *Electrophoresis* 19:998-1005, 1998). A variety of computer software programs to facilitate these comparisons are accessible via the Internet, such as Protein Prospector (Internet site: prospector.uscf.edu), MultiIdent (Internet site: www.expasy.ch/sprot/multiident.html), PeptideSearch (Internet site: www.mann.embl-heiedelberg.de...deSearch/FR_PeptideSearch Form.html), and ProFound (Internet site: www.chait-sgi.rockefeller.edu/cgi-bin/prot-id-frag.html). These programs allow the user to specify the cleavage agent and the molecular weights of the fragmented peptides within a designated tolerance. The programs compare these molecular weights to protein molecular weight information stored in databases to assist in determining the identity of the unknown protein. Accurate information concerning the number of fragmented peptides and the precise molecular weight of those peptides is required for accurate identification. Therefore, increasing the accuracy in determining the number of fragmented peptides and their molecular weight should result in enhanced likelihood of success in the identification of unknown proteins.

In addition, peptide digests of unknown proteins can be sequenced using tandem mass spectrometry (MS/MS), and the resulting sequence searched against databases (J.K. Eng, et al., *J. Am. Soc. Mass Spec.* 5:976-989 (1994); M. Mann and M. Wilm, *Anal. Chem.* 66:4390-4399

(1994); J.A. Taylor and R.S. Johnson. *Rapid Comm. Mass Spec.* 11:1067-1075 (1997)). Searching programs that can be used in this process exist on the Internet, such as Lutefisk 97 (Internet site: www.lsbc.com:70/Lutefisk97.html), and the Protein Prospector, Peptide Search and ProFound programs described above. Therefore, adding the sequence of a gene and its predicted protein sequence and peptide fragments to a sequence database can aid in the identification of unknown proteins using tandem mass spectrometry.

Thus, there also exists a need in the art for polypeptides suitable for use in peptide fragmentation studies, preferably, polypeptides that have altered expression in irritable bowel diseases, for use in molecular weight measurements, and for use in protein sequencing using tandem mass spectrometry.

## SUMMARY OF THE INVENTION

Using the T84 model, it was determined that certain polypeptides have altered (up-regulated or down-regulated) expression patterns in response to INF-gamma. Such molecules may have a role in gut barrier function and IBD and may be useful as potential therapeutic agents in the treatment of IBD and other gut pathologies. The invention aids in fulfilling these needs in the art by providing isolated nucleic acids and polypeptides encoded by these nucleic acids that have altered expression characteristics in the T84 gut barrier model. Particular embodiments of the invention are directed to isolated nucleic acid molecules comprising the DNA sequences of SEQ ID NOs:1-26 and isolated nucleic acid molecules encoding the amino acid sequences of SEQ ID NOs:27-38, as well as nucleic acid molecules complementary to these sequences. Both single-stranded and double-stranded RNA and DNA nucleic acid molecules are encompassed by the invention, as well as nucleic acid molecules that hybridize to a denatured, double-stranded DNA comprising all or a portion of SEQ ID NOs:1-26 and/or the DNA that encodes the amino acid sequences of SEQ ID NOs:27-38. Also encompassed are isolated nucleic acid molecules that are derived by *in vitro* mutagenesis from nucleic acid molecules comprising sequences of SEQ ID NOs:1-26, that are degenerate from nucleic acid molecules comprising sequences of SEQ ID NOs:1-26, and that are allelic variants of DNA of the invention. The invention also encompasses recombinant vectors that direct the expression of these nucleic acid molecules and host cells transformed or transfected with these vectors.

In another embodiment, the invention provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 80%, preferable 85%, more preferably 90%, optimally 95%, identical to a sequence of a polynucleotide selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:1-26 or a polynucleotide which is hybridizable to SEQ ID NO:1-26;

(b) a polynucleotide encoding a polypeptide fragment of a translation of SEQ ID NO: 1-26 or a polypeptide fragment encoded by the cDNA sequence which is hybridizable to SEQ ID NO:1-26;

(c) a polynucleotide encoding a polypeptide epitope of a translation of SEQ ID NO: 1-26 or a polypeptide epitope encoded by a cDNA sequence which is hybridizable to SEQ ID NO:1-26;

(e) a polynucleotide encoding a polypeptide of a translation of SEQ ID NO: 1-26, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:1-26;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:1-26;

(h) a polynucleotide which encodes a species homologue of a translation of SEQ ID NO: 1-26;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

In another embodiment, the invention provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 80%, preferable 85%, more preferably 90%, optimally 95%, identical to a sequence of a polynucleotide selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:1-10 or a polynucleotide which is hybridizable to SEQ ID NO:1-10;

(b) a polynucleotide encoding a polypeptide fragment of a translation of SEQ ID NO: 1-10 or a polypeptide fragment encoded by the cDNA sequence which is hybridizable to SEQ ID NO:1-10;

(c) a polynucleotide encoding a polypeptide epitope of a translation of SEQ ID NO: 1-10 or a polypeptide epitope encoded by a cDNA sequence which is hybridizable to SEQ ID NO:1-10;

(e) a polynucleotide encoding a polypeptide of a translation of SEQ ID NO: 1-10, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:1-10;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:1-10;

(h) a polynucleotide which encodes a species homologue of a translation of SEQ ID NO: 1-10;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

In another embodiment, the invention provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 80%, preferable 85%, more preferably 90%, optimally 95%, identical to a sequence of a polynucleotide selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:11-26 or a polynucleotide which is hybridizable to SEQ ID NO:11-26;

(b) a polynucleotide encoding a polypeptide fragment of a translation of SEQ ID NO: 11-26 or a polypeptide fragment encoded by the cDNA sequence which is hybridizable to SEQ ID NO:11-26;

(c) a polynucleotide encoding a polypeptide epitope of a translation of SEQ ID NO: 11-26 or a polypeptide epitope encoded by a cDNA sequence which is hybridizable to SEQ ID NO:11-26;

(e) a polynucleotide encoding a polypeptide of a translation of SEQ ID NO: 11-26, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:11-26;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:11-26;

(h) a polynucleotide which encodes a species homologue of a translation of SEQ ID NO: 11-26;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a secreted protein. In preferred embodiments, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a polypeptide chosen from the group consisting of:

(a) a polypeptide having the polypeptide sequence identified as a translation of SEQ ID NO: 1-26;

(b) a polypeptide having the polypeptide sequence of SEQ ID NO: 27-38; and

(c) a polypeptide encoded by the cDNA which is hybridizable to SEQ ID NO:1-26. Typically, the isolated nucleic acid molecule comprises the entire nucleotide sequence of SEQ ID NO:1-26 or a cDNA sequence which is hybridizable to SEQ ID NO:1-26. Typically the isolated nucleic acid molecule comprises sequential nucleotide deletions from portions of the nucleotide sequence encoding either the C-terminus or the N-terminus of the polypeptide.

In other aspects, the invention provides a recombinant vector comprising the isolated nucleic acid molecule and a method of making a recombinant host cell comprising the isolated nucleic acid molecule and the recombinant host cell produced by such a method.

In other preferred embodiments, the invention provides an isolated polypeptide having an amino acid sequence at least 80%, preferably at leasst 85%, more preferably at least 90% identical to the sequence of a polypeptide selected from the group consisting of:

(a) a polypeptide fragment of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26;

(b) a polypeptide having the sequence of SEQ ID NO: 27-38;

(c) a polypeptide domain of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26;

(d) a polypeptide epitope of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26;

(e) a secreted form of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26;

(f) a full length protein of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26;

(g) a variant of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26;

(h) an allelic variant of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26; and

(i) a species homologue of a polypeptide encoded by a polynucleotide of SEQ ID NO: 1-26.

In some embodiments, the full length polypeptide comprises sequential amino acid deletions from the C-terminus. In other embodiments, the mature polypeptide comprises sequential amino acid deletions from the C-terminus. Alternatively, the full length polypeptide can comprise sequential amino acid deletions from the N-terminus or the mature polypeptide can comprise sequential amino acid deletions from the N-terminus.

In other aspects. the invention provides an isolated antibody that binds specifically to the isolated polypeptide. a recombinant host cell that expresses the isolated polypeptide. and a method of making an isolated polypeptide comprising culturing the recombinant host cell under conditions such that said polypeptide is expressed; and recovering said polypeptide.

The invention in another embodiment is a method for preventing, treating, or ameliorating irritable bowel discorders, comprising administering to a mammalian subject a therapeutically effective amount of the isolated polypeptide or the isolated nucleic acid molecule. In another embodiment, the invention provides a method of diagnosing an irritable bowel disease or a susceptibility to irritable bowel disease in a subject comprising: determining the presence or absence of a mutation in the isolayed nucleic acid molecule, and diagnosing an irritable bowel disease or a susceptibility to irritable bowel disease based on the presence or absence of said mutation. Alternatively, irritable bowel disease or a susceptibility to irritable bowel disease in a subject can be diagnosed by a method comprising: determining the presence or amount of expression of the polypeptide in a biological sample; and diagnosing irritable bowel disease or a susceptibility to irritable bowel disease based on the presence or amount of expression of the polypeptide.

Typically, a binding partner of the polypeptide is identified by contacting the polypeptide with a binding partner; and determining whether the binding partner affects a physical property or an activity of the polypeptide. Typically activity in a biological assay of a secreted polypeptide is identified by expressing the polynucleotide of SEQ ID NO:1-26 in a cell; isolating the supernatant; detecting an activity in a biological assay; and identifying the polypeptide in the supernatant having the activity.

In addition, the invention encompasses methods of using the nucleic acids noted above to identify nucleic acids encoding proteins homologous to SEQ ID NOs:27-38; to identify human chromosomes that contain the nucleotide sequences of the invention; to map genes near the nucleotide sequences of the invention on human chromosomes; and to identify genes associated with certain diseases, syndromes, or other human conditions associated with human chromosomes containing sequences of the invention.

For example, four of the nucleotide sequences of the invention, IMX4 and IMX56, IMX21, and IMX44 are located on chromosomes 22, 22, 7, and 19, respectively (IMX4 and IMX56 are both located on chromosome 22). Thus, the above-named nucleotide sequences (IMX4, IMX56, IMX21, and IMX44) can be used to identify human chromosome numbers 22,

7, and 19; to map genes on human chromosome numbers 22, 7, and 19; and to identify genes associated with certain diseases, syndromes, or other human conditions associated with human chromosome numbers 22, 7, and 19.

The invention also encompasses the use of sense or antisense oligonucleotides from the nucleic acids of SEQ ID NOs:1-26 to inhibit the expression of the polynucleotides encoded by the nucleotide sequences of the invention.

The invention also encompasses isolated polypeptides and fragments thereof encoded by these nucleic acid molecules including soluble polypeptide portions of SEQ ID NOs:27-38. The invention further encompasses methods for the production of these polypeptides, including culturing a host cell under conditions promoting expression and recovering the polypeptide from the culture medium if it is secreted or from cultured cells if it is not secreted. Especially, the expression of these polypeptides in bacteria, yeast, plant, insect, and animal cells is encompassed by the invention.

In addition, the invention includes assays utilizing these polypeptides, to screen for potential inhibitors of activity associated with polypeptide counter-structure molecules, and methods of using these polypeptides as therapeutic agents for the treatment of diseases mediated by polypeptide counter-structure molecules. Further, methods of using these polypeptides in the design of inhibitors thereof are also an aspect of the invention.

The invention further includes a method for using these polypeptides as molecular weight markers that allow the estimation of the molecular weight of a protein or a fragmented protein, as well as a method for the visualization of the molecular weight markers of the invention thereof using electrophoresis. The invention further encompasses methods for using the polypeptides of the invention as markers for determining the isoelectric point of an unknown protein, as well as controls for establishing the extent of fragmentation of a protein.

Further encompassed by this invention are kits to aid in these determinations.

Further encompassed by this invention is the use of the IMX nucleic acid sequences, predicted amino acid sequences of the polypeptide or fragments thereof, or a combination of the predicted amino acid sequences of the polypeptides and fragments thereof for use in searching an electronic database to aid in the identification of sample nucleic acids and/or proteins.

The invention also encompasses IMX polypeptides and the use of these polypeptides as research reagents to further study gut epithelial barrier function and regulation and therapeutic reagents to treat inflammatory bowel disease and other gut pathologies.

Isolated polyclonal or monoclonal antibodies that bind to these polypeptides are also encompassed by the invention. in addition the use of these antibodies to aid in purifying IMX polypeptides.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 presents the nucleotide sequence of IMX 4 (SEQ ID NO: 1).

Figure 2 presents the nucleotide sequence of IMX 10 (SEQ ID NO: 14).

Figure 3 presents the nucleotide sequence of IMX 21 (SEQ ID NO: 15).

Figure 4 presents the nucleotide sequence of IMX 28 (SEQ ID NO: 16).

Figure 5 presents the nucleotide sequence of IMX 32 (SEQ ID NO: 17).

Figure 6 presents the nucleotide sequence of IMX 39 (SEQ ID NO: 20).

Figure 7 presents the nucleotide sequence of IMX 40 (SEQ ID NO: 7).

Figure 8 presents the nucleotide sequence of IMX 42 (SEQ ID NO: 8).

Figure 9 presents the nucleotide sequence of IMX 44 (SEQ ID NO: 23).

Figure 10 presents the nucleotide sequence of IMX 56 (SEQ ID NO: 10).

Figure 11 presents the amino acid sequence of IMX 4 (SEQ ID NO: 27).

Figure 12 presents the amino acid sequence of IMX 10 (SEQ ID NO: 28).

Figure 13 presents the amino acid sequence of IMX 21 (SEQ ID NO: 29).

Figure 14 presents the amino acid sequence of IMX 28 (SEQ ID NO: 30).

Figure 15 presents the amino acid sequence of IMX 32 (SEQ ID NO: 32).

Figure 16 presents the amino acid sequence of IMX 39 (SEQ ID NO: 33).

Figure 17 presents the amino acid sequence of IMX 40 (SEQ ID NO: 34).

Figure 18 presents the amino acid sequence of IMX 42 (SEQ ID NO: 35).

Figure 19 presents the amino acid sequence of IMX 44 (SEQ ID NO: 37).

Figure 20 presents the amino acid sequence of IMX 56 (SEQ ID NO: 38).

Figure 21 presents the nucleotide sequence of the 5' end of the clone (SEQ ID NO:13) matched to part of the human ApoL gene (AF019225).

Figure 22 presents comparison of the 5' end of the clone suggests it represents an alternative splice product to reported ApoL, i.., bases 1-168 match to 2 exons on PAC carrying the ApoL gene but are not included in the reported complete cDNA.

## DETAILED DESCRIPTION OF THE INVENTION

### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:1-26. For example, the polynucleotide can contain all or part of the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:1-26, or the complement thereof, or the cDNA. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of

hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M $NaH_2PO_4$; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 ug/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either

natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formulation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS – STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.).

The translated amino acid sequence, beginning with the methionine, is identified although other reading frames can also be easily translated using known molecular biology

techniques. The polypeptides produced by the translation of these alternative open reading frames are specifically contemplated by the present invention.

SEQ ID NO:1-26 and the translations of SEQ ID NO: 1-26 as well as SEQ ID NO:27-38 are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from the translations of SEQ ID NO:1-26 may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

The present invention also relates to the genes corresponding to SEQ ID NO:1-26, and translations of SEQ ID NO:1-26. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologues. Species homologues may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form,

and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

**Signal Sequences**

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. Therefore, from a deduced amino acid sequence, a signal sequence and mature sequence can be identified.

**Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

Further embodiments of the present invention include polynucleotides having at least 80% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to a sequence contained in SEQ ID NO:1-26. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity will encode a polypeptide identical to an amino acid sequence contained in the translations of SEQ ID NO:1-26.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference polypeptide, is intended that the amino acid sequence of the polypeptide is identical to the reference polypeptide except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the total length of the reference polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in

the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Further embodiments of the present invention include polypeptides having at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to an amino acid sequence contained in translations of SEQ ID NO: 1-26. Preferably, the above polypeptides should exhibit at least one biological activity of the protein.

In a preferred embodiment, polypeptides of the present invention include polypeptides having at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98%, or 99% similarity to an amino acid sequence contained in translations of SEQ ID NO:1-26 as well as the amino acid sequences of SEQ ID NO:27-38.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988 (1993) reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly,

Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1 a. They used random mutagenesis to generate over 3,500 individual IL-1 a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See Gayle et al., (1993), Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3.500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change. As the authors state, these two strategies, natural selection and genetic engineering, have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile;

replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

**Polynucleotide and Polypeptide Fragments**

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in that shown in SEQ ID NO:1-26. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in that shown in SEQ ID NO:1-26. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, and more nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-260, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, and so forth, to the end of SEQ ID NO:1-26. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably,

these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in the translations of SEQ ID NO:1-26 as well as SEQ ID NO:27-38. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, and so forth, to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50 or 60, amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of the translations of SEQ ID NO:1-26 as well as SEQ ID NO:27-38 falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

**Epitopes & Antibodies**

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

**Fusion Proteins**

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-0 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0 232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused

with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See. D. Bennett et al.. J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol. Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

**Vectors, Host Cells, and Protein Production**

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

**Nucleic Acid Molecules and Polypeptides of the Present Invention**

The nucleic acid molecules encompassed in the invention comprise nucleotide sequences SEQ ID NO: 1-26.

The amino acid sequences of the polypeptides encoded by the nucleotide sequences of the invention are given in SEQ ID NO:27-38.

The discovery of the nucleic acids of the invention enables the construction of expression vectors comprising nucleic acid sequences encoding polypeptides; host cells transfected or transformed with the expression vectors; isolated and purified biologically active polypeptides and fragments thereof; the use of the nucleic acids or oligonucleotides thereof as probes to identify nucleic acids encoding proteins.having amino acid sequences homologous to SEQ ID NOs: 27-38; the use of the nucleic acids or oligonucleotides thereof to identify human chromosomes, for example, 22, 7, and 19; the use of the nucleic acids or oligonucleotides

thereof to map genes on human chromosomes, for example, 22, 7, and 19; the use of the nucleic acids or oligonucleotides thereof to identify genes associated with certain diseases, syndromes or other human conditions associated with human chromosomes such as 22, 7, and 19; the use of single-stranded sense or antisense oligonucleotides from the nucleic acids to inhibit expression of polynucleotides encoded by the IMX sequences; the use of such polypeptides and soluble fragments as molecular weight markers; the use of such polypeptides and fragmented peptides as controls for peptide fragmentation, and kits comprising these reagents; the use of such polypeptides and fragments thereof to generate antibodies, and the use of antibodies to purify IMX polypeptides.

NUCLEIC ACID MOLECULES

In a particular embodiment, the invention relates to certain isolated nucleotide sequences that are free from contaminating endogenous material. A "nucleotide sequence" refers to a polynucleotide molecule in the form of a separate fragment or as a component of a larger nucleic acid construct. The nucleic acid molecule has been derived from DNA or RNA isolated at least once in substantially pure form and in a quantity or concentration enabling identification, manipulation, and recovery of its component nucleotide sequences by standard biochemical methods (such as those outlined in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd sed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989)). Such sequences are preferably provided and/or constructed in the form of an open reading frame uninterrupted by internal non-translated sequences, or introns, that are typically present in eukaryotic genes. Sequences of non-translated DNA can be present 5' or 3' from an open reading frame, where the same do not interfere with manipulation or expression of the coding region.

Nucleic acid molecules of the invention include DNA in both single-stranded and double-stranded form, as well as the RNA complement thereof. DNA includes, for example, cDNA, genomic DNA, chemically synthesized DNA, DNA amplified by PCR, and combinations thereof. Genomic DNA may be isolated by conventional techniques, *e.g.*, using the cDNA of SEQ ID NOs:1-26, or a suitable fragment thereof, as a probe.

The DNA molecules of the invention include full length genes as well as polynucleotides and fragments thereof. The full length gene may include the N-terminal signal peptide. Other embodiments include DNA encoding a soluble form, *e.g.*, encoding the extracellular domain of the protein, either with or without the signal peptide.

The nucleic acids of the invention are preferentially derived from human sources, but the invention includes those derived from non-human species, as well.

Preferred Sequences

The particularly preferred nucleotide sequences of the invention are SEQ ID NOs: 1-26, as set forth above. cDNA clones having the nucleotide sequence of SEQ ID NOs:1-26 were isolated as described in Example 1. The sequences of amino acids encoded by the DNA of SEQ ID NOs:1-26 are shown in SEQ ID NOs:27-38.

Additional Sequences

Due to the known degeneracy of the genetic code, wherein more than one codon can encode the same amino acid, a DNA sequence can vary from that shown in SEQ ID NOs:1-26, and still encode a polypeptide having the amino acid sequence of SEQ ID NOs:27-38. Such variant DNA sequences can result from silent mutations (e.g., occurring during PCR amplification), or can be the product of deliberate mutagenesis of a native sequence.

The invention thus provides additional isolated DNA sequences encoding polypeptides of the invention, selected from: (a) DNA comprising the nucleotide sequences of SEQ ID NOs: 1-26; (b) DNA encoding the polypeptides of SEQ ID NOs:27-38; (c) DNA capable of hybridization to a DNA of (a) or (b) under conditions of moderate stringency and which encodes polypeptides of the invention; (d) DNA capable of hybridization to a DNA of (a) or (b) under conditions of high stringency and which encodes polypeptides of the invention, and (e) DNA which is degenerate, as a result of the genetic code, to a DNA defined in (a), (b), (c), or (d) and which encode polypeptides of the invention. Of course, polypeptides encoded by such DNA sequences are encompassed by the invention.

As used herein, conditions of moderate stringency can be readily determined by those having ordinary skill in the art based on, for example, the length of the DNA. The basic conditions are set forth by Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2 ed. Vol. 1, pp. 1.101-264, Cold Spring Harbor Laboratory Press, (1989), and include use of a prewashing solution for the nitrocellulose filters 5X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0), hybridization conditions of about 50% formamide, 6X SSC at about 42°C (or other similar hybridization solution, such as Stark's solution, in about 50% formamide at about 42°C), and washing conditions of about 60°C, 0.5X SSC, 0.1% SDS. Conditions of high stringency can also be readily determined by the skilled artisan based on, for example, the length of the DNA. Generally, such conditions are defined as hybridization conditions as above, and with washing at approximately 68°C, 0.2X SSC, 0.1% SDS. The skilled artisan will recognize that the

temperature and wash solution salt concentration can be adjusted as necessary according to factors such as the length of the probe.

Also included as an embodiment of the invention is DNA encoding polypeptide fragments and polypeptides comprising inactivated N-glycosylation site(s), inactivated protease processing site(s), or conservative amino acid substitution(s), as described below.

In another embodiment, the nucleic acid molecules of the invention also comprise nucleotide sequences that are at least 80% identical to a native sequence. Also contemplated are embodiments in which a nucleic acid molecule comprises a sequence that is at least 90% identical, at least 95% identical, at least 98% identical, at least 99% identical, or at least 99.9% identical to a native sequence.

The percent identity may be determined by visual inspection and mathematical calculation. Alternatively, the percent identity of two nucleic acid sequences can be determined by comparing sequence information using the GAP computer program, version 6.0 described by Devereux et al. (*Nucl. Acids Res.* 12:387, 1984) and available from the University of Wisconsin Genetics Computer Group (UWGCG). The preferred default parameters for the GAP program include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) for nucleotides, and the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res. 14*:6745, 1986, as described by Schwartz and Dayhoff, eds., *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, pp. 353-358, 1979; (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps. Other programs used by one skilled in the art of sequence comparison may also be used.

The invention provides isolated nucleic acids useful in the production of polypeptides. Such polypeptides may be prepared by any of a number of conventional techniques. A DNA sequence encoding an IMX polypeptide or a desired fragment thereof may be subcloned into an expression vector for production of the polypeptide or fragment. The DNA sequence advantageously is fused to a sequence encoding a suitable leader or signal peptide. Alternatively, the desired fragment may be chemically synthesized using known techniques. DNA fragments also may be produced by restriction endonuclease digestion of a full length cloned DNA sequence, and isolated by electrophoresis on agarose gels. If necessary, oligonucleotides that reconstruct the 5' or 3' terminus to a desired point may be ligated to a DNA fragment generated by restriction enzyme digestion. Such oligonucleotides may additionally contain a restriction endonuclease cleavage site upstream of the desired coding sequence, and position an initiation codon (ATG) at the N-terminus of the coding sequence.

The well-known polymerase chain reaction (PCR) procedure also may be employed to isolate and amplify a DNA sequence encoding a desired protein fragment. Oligonucleotides that define the desired termini of the DNA fragment are employed as 5' and 3' primers. The oligonucleotides may additionally contain recognition sites for restriction endonucleases, to facilitate insertion of the amplified DNA fragment into an expression vector. PCR techniques are described in Saiki et al., *Science* 239:487 (1988); *Recombinant DNA Methodology*, Wu et al., eds., Academic Press, Inc., San Diego (1989), pp. 189-196; and *PCR Protocols: A Guide to Methods and Applications*, Innis et al., eds., Academic Press, Inc. (1990).

## POLYPEPTIDES AND FRAGMENTS THEREOF

The invention encompasses polypeptides and fragments thereof in various forms, including those that are naturally occurring or produced through various techniques such as procedures involving recombinant DNA technology. Such forms include, but are not limited to, derivatives, variants, and oligomers, as well as fusion proteins or fragments thereof.

### Polypeptides and Fragments Thereof

The polypeptides of the invention include full length proteins encoded by the nucleic acid sequences set forth above. Particularly preferred polypeptides comprise the amino acid sequences of SEQ ID NOs:27-38.

The polypeptides of the invention may include an N-terminal hydrophobic region that functions as a signal peptide, and may contain an extracellular domain, and may also contain a transmembrane region and a C-terminal cytoplasmic domain as well as a spacer region. Computer analysis may be used to predict the location of the signal peptide.

For example, the isolated polypeptides of SEQ ID NO:14 (IMX 28) and SEQ ID NO:19 (IMX 44) include an N-terminal hydrophobic region that functions as a signal peptide. Computer analysis predicts that the signal peptide corresponds to residues 1 to 17 of SEQ ID NO:14, and cleavage of the signal peptide of SEQ ID NO:14 results in a mature protein comprising amino acids 18 to 372. The signal peptide of IMX 44 corresponds to residues 1 to 37 of SEQ ID NO:19. The next most likely computer-predicted signal peptide cleavage sites (in descending order) would occur after amino acids 36, 26 and 27 of SEQ ID NO:19. Cleavage of the signal peptide at position 37 thus would yield a mature protein comprising amino acids 38 through 261 of SEQ ID NO:19.

The skilled artisan will recognize that the above-described boundaries of such regions of the polypeptide are approximate. To illustrate, the boundaries of the mature protein (which may

be predicted by using computer programs available for that purpose) may differ from those described above.

The polypeptides of the invention may be membrane bound or they may be secreted and, thus, soluble. Soluble polypeptides are capable of being secreted from the cells in which they are expressed. In general, soluble polypeptides may be identified (and distinguished from non-soluble membrane-bound counterparts) by separating intact cells which express the desired polypeptide from the culture medium, *e.g.*, by centrifugation, and assaying the medium (supernatant) for the presence of the desired polypeptide. The presence of polypeptide in the medium indicates that the polypeptide was secreted from the cells and thus is a soluble form of the protein.

In one embodiment, the soluble polypeptides and fragments thereof comprise all or part of the extracellular domain, but lack the transmembrane region that would cause retention of the polypeptide on a cell membrane. A soluble polypeptide may include the cytoplasmic domain, or a portion thereof, as long as the polypeptide is secreted from the cell in which it is produced.

In general, the use of soluble forms is advantageous for certain applications. Purification of the polypeptides from recombinant host cells is facilitated, since the soluble polypeptides are secreted from the cells. Further, soluble polypeptides are generally more suitable for intravenous administration.

The invention also provides polypeptides and fragments of the extracellular domain that retain a desired biological activity. Particular embodiments are directed to polypeptide fragments that retain the ability to bind the native cognate, substrate, or counter-structure ("binding partner"). Such a fragment may be a soluble polypeptide, as described above. In another embodiment, the polypeptides and fragments advantageously include regions that are conserved in the family as described above.

Also provided herein are polypeptide fragments comprising at least 20, or at least 30, contiguous amino acids of the sequence of SEQ ID NOs:27-38. Fragments derived from the cytoplasmic domain find use in studies of signal transduction, and in regulating cellular processes associated with transduction of biological signals. Polypeptide fragments also may be employed as immunogens, in generating antibodies.

Variants

Naturally occurring variants as well as derived variants of the polypeptides and fragments are provided herein.

Variants may exhibit amino acid sequences that are at least 80% identical. Also contemplated are embodiments in which a polypeptide or fragment comprises an amino acid sequence that is at least 90% identical, at least 95% identical, at least 98% identical, at least 99% identical, or at least 99.9% identical to the preferred polypeptide or fragment thereof. Percent identity may be determined by visual inspection and mathematical calculation. Alternatively, the percent identity of two protein sequences can be determined by comparing sequence information using the GAP computer program, based on the algorithm of Needleman and Wunsch (J. Mol. Bio. 48:443, 1970) and available from the University of Wisconsin Genetics Computer Group (UWGCG). The preferred default parameters for the GAP program include: (1) a scoring matrix, blosum62, as described by Henikoff and Henikoff (Proc. Natl. Acad. Sci. USA 89:10915, 1992); (2) a gap weight of 12; (3) a gap length weight of 4; and (4) no penalty for end gaps. Other programs used by one skilled in the art of sequence comparison may also be used.

The variants of the invention include, for example, those that result from alternate mRNA splicing events or from proteolytic cleavage. Alternate splicing of mRNA may, for example, yield a truncated but biologically active protein, such as a naturally occurring soluble form of the protein. Variations attributable to proteolysis include, for example, differences in the N- or C-termini upon expression in different types of host cells, due to proteolytic removal of one or more terminal amino acids from the protein (generally from 1-5 terminal amino acids). Proteins in which differences in amino acid sequence are attributable to genetic polymorphism (allelic variation among individuals producing the protein) are also contemplated herein.

Additional variants within the scope of the invention include polypeptides that may be modified to create derivatives thereof by forming covalent or aggregative conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives may be prepared by linking the chemical moieties to functional groups on amino acid side chains or at the N-terminus or C-terminus of a polypeptide. Conjugates comprising diagnostic (detectable) or therapeutic agents attached thereto are contemplated herein, as discussed in more detail below.

Other derivatives include covalent or aggregative conjugates of the polypeptides with other proteins or polypeptides, such as by synthesis in recombinant culture as N-terminal or C-terminal fusions. Examples of fusion proteins are discussed below in connection with oligomers. Further, fusion proteins can comprise peptides added to facilitate purification and identification. Such peptides include, for example, poly-His or the antigenic identification peptides described in U.S. Patent No. 5,011,912 and in Hopp et al., *Bio/Technology* 6:1204,

1988. One such peptide is the FLAG® peptide. Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys, which is highly antigenic and provides an epitope reversibly bound by a specific monoclonal antibody, enabling rapid assay and facile purification of expressed recombinant protein. A murine hybridoma designated 4E11 produces a monoclonal antibody that binds the FLAG® peptide in the presence of certain divalent metal cations, as described in U.S. Patent 5,011,912, hereby incorporated by reference. The 4E11 hybridoma cell line has been deposited with the American Type Culture Collection under accession no. HB 9259. Monoclonal antibodies that bind the FLAG® peptide are available from Eastman Kodak Co., Scientific Imaging Systems Division, New Haven, Connecticut.

Among the polypeptides provided herein are variants of native polypeptides that retain the native biological activity or the substantial equivalent thereof. One example is a variant that binds with essentially the same binding affinity as does the native form. Binding affinity can be measured by conventional procedures, *e.g.*, as described in U.S. Patent No. 5,512,457 and as set forth below.

Variants include polypeptides that are substantially homologous to the native form, but which have an amino acid sequence different from that of the native form because of one or more deletions, insertions or substitutions. Particular embodiments include, but are not limited to, polypeptides that comprise from one to ten deletions, insertions or substitutions of amino acid residues, when compared to a native sequence.

A given amino acid may be replaced, for example, by a residue having similar physiochemical characteristics. Examples of such conservative substitutions include substitution of one aliphatic residue for another, such as Ile, Val, Leu, or Ala for one another; substitutions of one polar residue for another, such as between Lys and Arg, Glu and Asp, or Gln and Asn; or substitutions of one aromatic residue for another, such as Phe, Trp, or Tyr for one another. Other conservative substitutions, *e.g.*, involving substitutions of entire regions having similar hydrophobicity characteristics, are well known.

Similarly, the DNAs of the invention include variants that differ from a native DNA sequence because of one or more deletions, insertions or substitutions, but that encode a biologically active polypeptide.

The invention further includes polypeptides of the invention with or without associated native-pattern glycosylation. Polypeptides expressed in yeast or mammalian expression systems (*e.g.*, COS-1 or COS-7 cells) can be similar to or significantly different from a native polypeptide in molecular weight and glycosylation pattern, depending upon the choice of expression system. Expression of polypeptides of the invention in bacterial expression systems.

such as *E. coli*, provides non-glycosylated molecules. Further, a given preparation may include multiple differentially glycosylated species of the protein. Glycosyl groups can be removed through conventional methods, in particular those utilizing glycopeptidase. In general, glycosylated polypeptides of the invention can be incubated with a molar excess of glycopeptidase (Boehringer Mannheim).

Correspondingly, similar DNA constructs that encode various additions or substitutions of amino acid residues or sequences, or deletions of terminal or internal residues or sequences are encompassed by the invention. For example, N-glycosylation sites in the polypeptide extracellular domain can be modified to preclude glycosylation, allowing expression of a reduced carbohydrate analog in mammalian and yeast expression systems. N-glycosylation sites in eukaryotic polypeptides are characterized by an amino acid triplet Asn-X-Y, wherein X is any amino acid except Pro and Y is Ser or Thr. Appropriate substitutions, additions, or deletions to the nucleotide sequence encoding these triplets will result in prevention of attachment of carbohydrate residues at the Asn side chain. Alteration of a single nucleotide, chosen so that Asn is replaced by a different amino acid, for example, is sufficient to inactivate an N-glycosylation site. Alternatively, the Ser or Thr can by replaced with another amino acid, such as Ala. Known procedures for inactivating N-glycosylation sites in proteins include those described in U.S. Patent 5,071,972 and EP 276,846, hereby incorporated by reference.

In another example of variants, sequences encoding Cys residues that are not essential for biological activity can be altered to cause the Cys residues to be deleted or replaced with other amino acids, preventing formation of incorrect intramolecular disulfide bridges upon folding or renaturation.

Other variants are prepared by modification of adjacent dibasic amino acid residues, to enhance expression in yeast systems in which KEX2 protease activity is present. EP 212,914 discloses the use of site-specific mutagenesis to inactivate KEX2 protease processing sites in a protein. KEX2 protease processing sites are inactivated by deleting, adding or substituting residues to alter Arg-Arg, Arg-Lys, and Lys-Arg pairs to eliminate the occurrence of these adjacent basic residues. Lys-Lys pairings are considerably less susceptible to KEX2 cleavage, and conversion of Arg-Lys or Lys-Arg to Lys-Lys represents a conservative and preferred approach to inactivating KEX2 sites.

Oligomers

Encompassed by the invention are oligomers or fusion proteins that contain IMX polypeptides. Such oligomers may be in the form of covalently-linked or non-covalently-linked multimers, including dimers. trimers, or higher oligomers. As noted above, preferred polypeptides are soluble and thus these oligomers may comprise soluble polypeptides. In one aspect of the invention, the oligomers maintain the binding ability of the polypeptide components and provide therefor, bivalent, trivalent, etc., binding sites.

One embodiment of the invention is directed to oligomers comprising multiple polypeptides joined *via* covalent or non-covalent interactions between peptide moieties fused to the polypeptides. Such peptides may be peptide linkers (spacers), or peptides that have the property of promoting oligomerization. Leucine zippers and certain polypeptides derived from antibodies are among the peptides that can promote oligomerization of the polypeptides attached thereto, as described in more detail below.

Immunoglobulin-based Oligomers

As one alternative, an oligomer is prepared using polypeptides derived from immunoglobulins. Preparation of fusion proteins comprising certain heterologous polypeptides fused to various portions of antibody-derived polypeptides (including the Fc domain) has been described, *e.g.*, by Ashkenazi et al. (*PNAS USA* 88:10535, 1991); Byrn et al. (*Nature* 344:677, 1990); and Hollenbaugh and Aruffo ("Construction of Immunoglobulin Fusion Proteins", in *Current Protocols in Immunology*, Suppl. 4, pages 10.19.1 - 10.19.11, 1992).

One embodiment of the present invention is directed to a dimer comprising two fusion proteins created by fusing a polypeptide of the invention to an Fc polypeptide derived from an antibody. A gene fusion encoding the polypeptide/Fc fusion protein is inserted into an appropriate expression vector. Polypeptide/Fc fusion proteins are expressed in host cells transformed with the recombinant expression vector, and allowed to assemble much like antibody molecules, whereupon interchain disulfide bonds form between the Fc moieties to yield divalent molecules.

The term "Fc polypeptide" as used herein includes native and mutein forms of polypeptides comprising from the Fc region of an antibody. Truncated forms of such polypeptides containing the hinge region that promotes dimerization are also included. Preferred polypeptides comprise an Fc polypeptide derived from a human IgG1 antibody comprising any or all of the CH domains of the Fc region.

One suitable Fc polypeptide, described in PCT application WO 93/10151, hereby incorporated by reference, is a single chain polypeptide extending from the N-terminal hinge

region to the native C-terminus of the Fc region of a human IgG1 antibody. Another useful Fc polypeptide is the Fc mutein described in U.S. Patent 5,457,035 and in Baum et al., (*EMBO J.* 13:3992-4001, 1994) incorporated herein by reference. The amino acid sequence of this mutein is identical to that of the native Fc sequence presented in WO 93/10151, except that amino acid 19 has been changed from Leu to Ala, amino acid 20 has been changed from Leu to Glu, and amino acid 22 has been changed from Gly to Ala. The mutein exhibits reduced affinity for Fc receptors.

The above-described fusion proteins comprising Fc moieties (and oligomers formed therefrom) offer the advantage of facile purification by affinity chromatography over Protein A or Protein G columns.

In other embodiments, the polypeptides of the invention may be substituted for the variable portion of an antibody heavy or light chain. If fusion proteins are made with both heavy and light chains of an antibody, it is possible to form an oligomer with as many as four soluble regions of the proteins of the invention.

Peptide-linker Based Oligomers

Alternatively, the oligomer is a fusion protein comprising multiple polypeptides, with or without peptide linkers (spacer peptides). Among the suitable peptide linkers are those described in U.S. Patents 4,751,180 and 4,935,233, which are hereby incorporated by reference. A DNA sequence encoding a desired peptide linker may be inserted between, and in the same reading frame as, the DNA sequences of the invention, using any suitable conventional technique. For example, a chemically synthesized oligonucleotide encoding the linker may be ligated between the sequences. In particular embodiments, a fusion protein comprises from two to four soluble IMX polypeptides, separated by peptide linkers.

Leucine-Zippers

Another method for preparing the oligomers of the invention involves use of a leucine zipper. Leucine zipper domains are peptides that promote oligomerization of the proteins in which they are found. Leucine zippers were originally identified in several DNA-binding proteins (Landschulz et al., *Science* 240:1759, 1988), and have since been found in a variety of different proteins. Among the known leucine zippers are naturally occurring peptides and derivatives thereof that dimerize or trimerize.

Examples of leucine zipper domains suitable for producing soluble oligomeric proteins are described in PCT application WO 94/10308, and the leucine zipper derived from lung

surfactant protein D (SPD) described in Hoppe et al. (*FEBS Letters* 344:191, 1994), hereby incorporated by reference. The use of a modified leucine zipper that allows for stable trimerization of a heterologous protein fused thereto is described in Fanslow et al. (*Semin. Immunol.* 6:267-278, 1994). Recombinant fusion proteins comprising a soluble polypeptide fused to a leucine zipper peptide are expressed in suitable host cells, and the soluble oligomer that forms is recovered from the culture supernatant.

In particular embodiments, leucine residues in a leucine zipper moiety are replaced by isoleucine residues. Such peptides comprising isoleucine may be referred to as "isoleucine zippers" but are encompassed by the term "leucine zippers" as employed herein.

## PRODUCTION OF POLYPEPTIDES AND FRAGMENTS THEREOF

Expression, isolation and purification of the polypeptides and fragments of the invention may be accomplished by any suitable technique, including but not limited to the following:

### Expression Systems

The present invention also provides recombinant cloning and expression vectors containing DNA, as well as host cell containing the recombinant vectors. Expression vectors comprising DNA may be used to prepare the polypeptides or fragments of the invention encoded by the DNA. A method for producing polypeptides comprises culturing host cells transformed with a recombinant expression vector encoding the polypeptide, under conditions that promote expression of the polypeptide, then recovering the expressed polypeptides from the culture. The skilled artisan will recognize that the procedure for purifying the expressed polypeptides will vary according to such factors as the type of host cells employed, and whether the polypeptide is membrane-bound or a soluble form that is secreted from the host cell.

Any suitable expression system may be employed. The vectors include a DNA encoding a polypeptide or fragment of the invention, operably linked to suitable transcriptional or translational regulatory nucleotide sequences, such as those derived from a mammalian, microbial, viral, or insect gene. Examples of regulatory sequences include transcriptional promoters, operators, or enhancers, an mRNA ribosomal binding site, and appropriate sequences which control transcription and translation initiation and termination. Nucleotide sequences are operably linked when the regulatory sequence functionally relates to the DNA sequence. Thus, a promoter nucleotide sequence is operably linked to a DNA sequence if the promoter nucleotide sequence controls the transcription of the DNA sequence. An origin of replication that confers

the ability to replicate in the desired host cells. and a selection gene by which transformants are identified. are generally incorporated into the expression vector.

In addition, a sequence encoding an appropriate signal peptide (native or heterologous) can be incorporated into expression vectors. A DNA sequence for a signal peptide (secretory leader) may be fused in frame to the nucleic acid sequence of the invention so that the DNA is initially transcribed, and the mRNA translated, into a fusion protein comprising the signal peptide. A signal peptide that is functional in the intended host cells promotes extracellular secretion of the polypeptide. The signal peptide is cleaved from the polypeptide upon secretion of polypeptide from the cell.

The skilled artisan will also recognize that the position(s) at which the signal peptide is cleaved may differ from that predicted by computer program, and may vary according to such factors as the type of host cells employed in expressing a recombinant polypeptide. A protein preparation may include a mixture of protein molecules having different N-terminal amino acids, resulting from cleavage of the signal peptide at more than one site.

Suitable host cells for expression of polypeptides include prokaryotes, yeast or higher eukaryotic cells. Mammalian or insect cells are generally preferred for use as host cells. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are described, for example, in Pouwels et al. *Cloning Vectors: A Laboratory Manual*, Elsevier, New York, (1985). Cell-free translation systems could also be employed to produce polypeptides using RNAs derived from DNA constructs disclosed herein.

Prokaryotic Systems

Prokaryotes include gram-negative or gram-positive organisms. Suitable prokaryotic host cells for transformation include, for example, *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium*, and various other species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. In a prokaryotic host cell, such as *E. coli*, a polypeptide may include an N-terminal methionine residue to facilitate expression of the recombinant polypeptide in the prokaryotic host cell. The N-terminal Met may be cleaved from the expressed recombinant polypeptide.

Expression vectors for use in prokaryotic host cells generally comprise one or more phenotypic selectable marker genes. A phenotypic selectable marker gene is, for example, a gene encoding a protein that confers antibiotic resistance or that supplies an autotrophic requirement. Examples of useful expression vectors for prokaryotic host cells include those derived from commercially available plasmids such as the cloning vector pBR322 (ATCC

37017). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides simple means for identifying transformed cells. An appropriate promoter and a DNA sequence are inserted into the pBR322 vector. Other commercially available vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotec, Madison, WI, USA).

Promoter sequences commonly used for recombinant prokaryotic host cell expression vectors include β-lactamase (penicillinase), lactose promoter system (Chang et al., *Nature 275*:615, 1978; and Goeddel et al., *Nature 281*:544, 1979), tryptophan (trp) promoter system (Goeddel et al., *Nucl. Acids Res. 8*:4057, 1980; and EP-A-36776) and tac promoter (Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, p. 412, 1982). A particularly useful prokaryotic host cell expression system employs a phage $\lambda P_L$ promoter and a cI857ts thermolabile repressor sequence. Plasmid vectors available from the American Type Culture Collection which incorporate derivatives of the $\lambda P_L$ promoter include plasmid pHUB2 (resident in *E. coli* strain JMB9, ATCC 37092) and pPLc28 (resident in *E. coli* RR1, ATCC 53082).

Yeast Systems

Alternatively, the polypeptides may be expressed in yeast host cells, preferably from the *Saccharomyces* genus (*e.g., S. cerevisiae*). Other genera of yeast, such as *Pichia* or *Kluyveromyces*, may also be employed. Yeast vectors will often contain an origin of replication sequence from a 2μ yeast plasmid, an autonomously replicating sequence (ARS), a promoter region, sequences for polyadenylation, sequences for transcription termination, and a selectable marker gene. Suitable promoter sequences for yeast vectors include, among others, promoters for metallothionein, 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem. 255*:2073, 1980) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg. 7*:149, 1968; and Holland et al., *Biochem. 17*:4900, 1978), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phospho-glucose isomerase, and glucokinase. Other suitable vectors and promoters for use in yeast expression are further described in Hitzeman, EPA-73,657. Another alternative is the glucose-repressible ADH2 promoter described by Russell et al. (*J. Biol. Chem. 258*:2674, 1982) and Beier et al. (*Nature 300*:724, 1982). Shuttle vectors replicable in both yeast and *E. coli* may be constructed by inserting DNA sequences from pBR322 for selection and replication in *E. coli* ($Amp^r$ gene and origin of replication) into the above-described yeast vectors.

The yeast α-factor leader sequence may be employed to direct secretion of the polypeptide. The α-factor leader sequence is often inserted between the promoter sequence and the structural gene sequence. *See*, e.g., Kurjan et al., *Cell 30*:933, 1982 and Bitter et al., *Proc. Natl. Acad. Sci. USA 81*:5330, 1984. Other leader sequences suitable for facilitating secretion of recombinant polypeptides from yeast hosts are known to those of skill in the art. A leader sequence may be modified near its 3' end to contain one or more restriction sites. This will facilitate fusion of the leader sequence to the structural gene.

Yeast transformation protocols are known to those of skill in the art. One such protocol is described by Hinnen et al., *Proc. Natl. Acad. Sci. USA 75*:1929, 1978. The Hinnen et al. protocol selects for $Trp^+$ transformants in a selective medium, wherein the selective medium consists of 0.67% yeast nitrogen base, 0.5% casamino acids, 2% glucose, 10 mg/ml adenine and 20 mg/ml uracil.

Yeast host cells transformed by vectors containing an ADH2 promoter sequence may be grown for inducing expression in a "rich" medium. An example of a rich medium is one consisting of 1% yeast extract, 2% peptone, and 1% glucose supplemented with 80 mg/ml adenine and 80 mg/ml uracil. Derepression of the ADH2 promoter occurs when glucose is exhausted from the medium.

Mammalian or Insect Systems

Mammalian or insect host cell culture systems also may be employed to express recombinant polypeptides. Baculovirus systems for production of heterologous proteins in insect cells are reviewed by Luckow and Summers, *Bio/Technology* 6:47 (1988). Established cell lines of mammalian origin also may be employed. Examples of suitable mammalian host cell lines include the COS-7 line of monkey kidney cells (ATCC CRL 1651) (Gluzman et al., *Cell 23*:175, 1981), L cells, C127 cells, 3T3 cells (ATCC CCL 163), Chinese hamster ovary (CHO) cells, HeLa cells, and BHK (ATCC CRL 10) cell lines, and the CV1/EBNA cell line derived from the African green monkey kidney cell line CV1 (ATCC CCL 70) as described by McMahan et al. (*EMBO J.* 10: 2821, 1991).

Established methods for introducing DNA into mammalian cells have been described (Kaufman, R.J., *Large Scale Mammalian Cell Culture*, 1990, pp. 15-69). Additional protocols using commercially available reagents, such as Lipofectamine lipid reagent (Gibco/BRL) or Lipofectamine-Plus lipid reagent, can be used to transfect cells (Felgner et al., *Proc. Natl. Acad. Sci. USA 84*:7413-7417, 1987). In addition, electroporation can be used to transfect mammalian cells using conventional procedures, such as those in Sambrook et al. (*Molecular Cloning: A*

*Laboratory Manual*, 2 ed. Vol. 1-3, Cold Spring Harbor Laboratory Press, 1989). Selection of stable transformants can be performed using methods known in the art, such as, for example, resistance to cytotoxic drugs. Kaufman et al., *Meth. in Enzymology 185*:487-511, 1990, describes several selection schemes, such as dihydrofolate reductase (DHFR) resistance. A suitable host strain for DHFR selection can be CHO strain DX-B11, which is deficient in DHFR (Urlaub and Chasin, *Proc. Natl. Acad. Sci. USA 77*:4216-4220, 1980). A plasmid expressing the DHFR cDNA can be introduced into strain DX-B11, and only cells that contain the plasmid can grow in the appropriate selective media. Other examples of selectable markers that can be incorporated into an expression vector include cDNAs conferring resistance to antibiotics, such as G418 and hygromycin B. Cells harboring the vector can be selected on the basis of resistance to these compounds.

Transcriptional and translational control sequences for mammalian host cell expression vectors can be excised from viral genomes. Commonly used promoter sequences and enhancer sequences are derived from polyoma virus, adenovirus 2, simian virus 40 (SV40), and human cytomegalovirus. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites can be used to provide other genetic elements for expression of a structural gene sequence in a mammalian host cell. Viral early and late promoters are particularly useful because both are easily obtained from a viral genome as a fragment, which can also contain a viral origin of replication (Fiers et al., *Nature 273*:113, 1978; Kaufman, *Meth. in Enzymology*, 1990). Smaller or larger SV40 fragments can also be used, provided the approximately 250 bp sequence extending from the *Hind* III site toward the *Bgl* I site located in the SV40 viral origin of replication site is included.

Additional control sequences shown to improve expression of heterologous genes from mammalian expression vectors include such elements as the expression augmenting sequence element (EASE) derived from CHO cells (Morris et al., *Animal Cell Technology*, 1997, pp. 529-534 and PCT Application WO 97/25420) and the tripartite leader (TPL) and VA gene RNAs from Adenovirus 2 (Gingeras et al., *J. Biol. Chem. 257*:13475-13491, 1982). The internal ribosome entry site (IRES) sequences of viral origin allows dicistronic mRNAs to be translated efficiently (Oh and Sarnow, *Current Opinion in Genetics and Development 3*:295-300, 1993; Ramesh et al., *Nucleic Acids Research 24*:2697-2700, 1996). Expression of a heterologous cDNA as part of a dicistronic mRNA followed by the gene for a selectable marker (e.g. DHFR) has been shown to improve transfectability of the host and expression of the heterologous cDNA (Kaufman, *Meth. in Enzymology*, 1990). Exemplary expression vectors that employ dicistronic

mRNAs are pTR-DC/GFP described by Mosser et al., *Biotechniques* 22:150-161, 1997, and p2A5I described by Morris et al., *Animal Cell Technology*, 1997, pp. 529-534.

A useful high expression vector, pCAVNOT, has been described by Mosley et al., *Cell 59*:335-348, 1989. Other expression vectors for use in mammalian host cells can be constructed as disclosed by Okayama and Berg (*Mol. Cell. Biol. 3*:280, 1983). A useful system for stable high level expression of mammalian cDNAs in C127 murine mammary epithelial cells can be constructed substantially as described by Cosman et al. (*Mol. Immunol. 23*:935, 1986). A useful high expression vector, PMLSV N1/N4, described by Cosman et al., *Nature 312*:768, 1984, has been deposited as ATCC 39890. Additional useful mammalian expression vectors are described in EP-A-0367566, and in WO 91/18982, incorporated by reference herein. In yet another alternative, the vectors can be derived from retroviruses.

Another useful expression vector, pFLAG®, can be used. FLAG® technology is centered on the fusion of a low molecular weight (1kD), hydrophilic, FLAG® marker peptide to the N-terminus of a recombinant protein expressed by pFLAG® expression vectors.

Regarding signal peptides that may be employed, the native signal peptide may be replaced by a heterologous signal peptide or leader sequence, if desired. The choice of signal peptide or leader may depend on factors such as the type of host cells in which the recombinant polypeptide is to be produced. To illustrate, examples of heterologous signal peptides that are functional in mammalian host cells include the signal sequence for interleukin-7 (IL-7) described in United States Patent 4,965,195, the signal sequence for interleukin-2 receptor described in Cosman et al., *Nature* 312:768 (1984); the interleukin-4 receptor signal peptide described in EP 367,566; the type I interleukin-1 receptor signal peptide described in U.S. Patent 4,968,607; and the type II interleukin-1 receptor signal peptide described in EP 460,846.

Purification

The invention also includes methods of isolating and purifying the polypeptides and fragments thereof.

Isolation and Purification

The "isolated" polypeptides or fragments thereof encompassed by this invention are polypeptides or fragments that are not in an environment identical to an environment in which it or they can be found in nature. The "purified" polypeptides or fragments thereof encompassed by this invention are essentially free of association with other proteins or polypeptides, for example, as a purification product of recombinant expression systems such as those described above or as a purified product from a non-recombinant source such as naturally occurring cells and/or tissues.

In one preferred embodiment, the purification of recombinant polypeptides or fragments can be accomplished using fusions of polypeptides or fragments of the invention to another polypeptide to aid in the purification of polypeptides or fragments of the invention. Such fusion partners can include the poly-His or other antigenic identification peptides described above as well as the Fc moieties described previously.

With respect to any type of host cell, as is known to the skilled artisan, procedures for purifying a recombinant polypeptide or fragment will vary according to such factors as the type of host cells employed and whether or not the recombinant polypeptide or fragment is secreted into the culture medium.

In general, the recombinant polypeptide or fragment can be isolated from the host cells if not secreted, or from the medium or supernatant if soluble and secreted, followed by one or more concentration, salting-out, ion exchange, hydrophobic interaction, affinity purification or size exclusion chromatography steps. As to specific ways to accomplish these steps, the culture medium first can be concentrated using a commercially available protein concentration filter, for example, an Amicon or Millipore Pellicon ultrafiltration unit. Following the concentration step, the concentrate can be applied to a purification matrix such as a gel filtration medium. Alternatively, an anion exchange resin can be employed, for example, a matrix or substrate having pendant diethylaminoethyl (DEAE) groups. The matrices can be acrylamide, agarose, dextran, cellulose or other types commonly employed in protein purification. Alternatively, a cation exchange step can be employed. Suitable cation exchangers include various insoluble matrices comprising sulfopropyl or carboxymethyl groups. In addition, a chromatofocusing step can be employed. Alternatively, a hydrophobic interaction chromatography step can be employed. Suitable matrices can be phenyl or octyl moieties bound to resins. In addition, affinity chromatography with a matrix which selectively binds the recombinant protein can be employed. Examples of such resins employed are lectin columns, dye columns, and metal-chelating columns. Finally, one or more reverse-phase high performance liquid chromatography

(RP-HPLC) steps employing hydrophobic RP-HPLC media. (e.g., silica gel or polymer resin having pendant methyl, octyl, octyldecyl or other aliphatic groups) can be employed to further purify the polypeptides. Some or all of the foregoing purification steps, in various combinations, are well known and can be employed to provide an isolated and purified recombinant protein.

It is also possible to utilize an affinity column comprising a polypeptide-binding protein of the invention, such as a monoclonal antibody generated against polypeptides of the invention, to affinity-purify expressed polypeptides. These polypeptides can be removed from an affinity column using conventional techniques, e.g., in a high salt elution buffer and then dialyzed into a lower salt buffer for use or by changing pH or other components depending on the affinity matrix utilized, or be competitively removed using the naturally occurring substrate of the affinity moiety, such as a polypeptide derived from the invention.

In this aspect of the invention, polypeptide-binding proteins, such as the anti-polypeptide antibodies of the invention or other proteins that may interact with the polypeptide of the invention, can be bound to a solid phase support such as a column chromatography matrix or a similar substrate suitable for identifying, separating, or purifying cells that express polypeptides of the invention on their surface. Adherence of polypeptide-binding proteins of the invention to a solid phase contacting surface can be accomplished by any means. For example, magnetic microspheres can be coated with these polypeptide-binding proteins and held in the incubation vessel through a magnetic field. Suspensions of cell mixtures are contacted with the solid phase that has such polypeptide-binding proteins thereon. Cells having polypeptides of the invention on their surface bind to the fixed polypeptide-binding protein and unbound cells then are washed away. This affinity-binding method is useful for purifying, screening, or separating such polypeptide-expressing cells from solution. Methods of releasing positively selected cells from the solid phase are known in the art and encompass, for example, the use of enzymes. Such enzymes are preferably non-toxic and non-injurious to the cells and are preferably directed to cleaving the cell-surface binding partner.

Alternatively, mixtures of cells suspected of containing polypeptide-expressing cells of the invention first can be incubated with a biotinylated polypeptide-binding protein of the invention. Incubation periods are typically at least one hour in duration to ensure sufficient binding to polypeptides of the invention. The resulting mixture then is passed through a column packed with avidin-coated beads, whereby the high affinity of biotin for avidin provides the binding of the polypeptide-binding cells to the beads. Use of avidin-coated beads is known in

the art. See Berenson, et al. *J. Cell. Biochem.*, 10D:239 (1986). Wash of unbound material and the release of the bound cells is performed using conventional methods.

The desired degree of purity depends on the intended use of the protein. A relatively high degree of purity is desired when the polypeptide is to be administered *in vivo*, for example. In such a case, the polypeptides are purified such that no protein bands corresponding to other proteins are detectable upon analysis by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). It will be recognized by one skilled in the pertinent field that multiple bands corresponding to the polypeptide may be visualized by SDS-PAGE, due to differential glycosylation, differential post-translational processing, and the like. Most preferably, the polypeptide of the invention is purified to substantial homogeneity, as indicated by a single protein band upon analysis by SDS-PAGE. The protein band may be visualized by silver staining, Coomassie blue staining, or (if the protein is radiolabeled) by autoradiography.

Assays

The purified polypeptides of the invention (including proteins, polypeptides, fragments, variants, oligomers, and other forms) may be tested for the ability to bind a cognate, ligand, receptor, substrate, or counter-structure and the like ("binding partner") in any suitable assay, such as a conventional binding assay. To illustrate, the polypeptide may be labeled with a detectable reagent (e.g., a radionuclide, chromophore, enzyme that catalyzes a colorimetric or fluorometric reaction, and the like). The labeled polypeptide is contacted with cells expressing the binding partner. The cells then are washed to remove unbound labeled polypeptide, and the presence of cell-bound label is determined by a suitable technique, chosen according to the nature of the label.

One example of a binding assay procedure is as follows. A recombinant expression vector containing the binding partner cDNA is constructed using methods known in the art. CV1-EBNA-1 cells in 10 $cm^2$ dishes are transfected with the recombinant expression vector. CV-1/EBNA-1 cells (ATCC CRL 10478) constitutively express EBV nuclear antigen-1 driven from the CMV immediate-early enhancer/promoter. CV1-EBNA-1 was derived from the African Green Monkey kidney cell line CV-1 (ATCC CCL 70), as described by McMahan et al. (*EMBO J.* 10:2821, 1991).

The transfected cells are cultured for 24 hours, and the cells in each dish then are split into a 24-well plate. After culturing an additional 48 hours, the transfected cells (about 4 x $10^4$ cells/well) are washed with BM-NFDM, which is binding medium (RPMI 1640 containing 25 mg/ml bovine serum albumin, 2 mg/ml sodium azide, 20 mM Hepes pH 7.2) to which 50 mg/ml

nonfat dry milk has been added. The cells then are incubated for 1 hour at 37°C with various concentrations of, for example, a soluble polypeptide/Fc fusion protein made as set forth above. Cells then are washed and incubated with a constant saturating concentration of a $^{125}$I-mouse anti-human IgG in binding medium, with gentle agitation for 1 hour at 37°C. After extensive washing, cells are released *via* trypsinization.

The mouse anti-human IgG employed above is directed against the Fc region of human IgG and can be obtained from Jackson Immunoresearch Laboratories, Inc., West Grove, PA. The antibody is radioiodinated using the standard chloramine-T method. The antibody will bind to the Fc portion of any polypeptide/Fc protein that has bound to the cells. In all assays, non-specific binding of $^{125}$I-antibody is assayed in the absence of the Fc fusion protein/Fc, as well as in the presence of the Fc fusion protein and a 200-fold molar excess of unlabeled mouse anti-human IgG antibody.

Cell-bound $^{125}$I-antibody is quantified on a Packard Autogamma counter. Affinity calculations (Scatchard, *Ann. N.Y. Acad. Sci.* 51:660, 1949) are generated on RS/1 (BBN Software, Boston, MA) run on a Microvax computer.

Another type of suitable binding assay is a competitive binding assay. To illustrate, biological activity of a variant may be determined by assaying for the variant's ability to compete with the native protein for binding to the binding partner.

Competitive binding assays can be performed by conventional methodology. Reagents that may be employed in competitive binding assays include radiolabeled IMX polypeptides and intact cells expressing the IMX polypeptide (endogenous or recombinant) on the cell surface. For example, a radiolabeled soluble IMX polypeptide fragment can be used to compete with a soluble IMX polypeptide variant for binding to the cell surface binding partner. Instead of intact cells, one could substitute a soluble binding partner/Fc fusion protein bound to a solid phase through the interaction of Protein A or Protein G (on the solid phase) with the Fc moiety. Chromatography columns that contain Protein A and Protein G include those available from Pharmacia Biotech, Inc., Piscataway, NJ.

Another type of competitive binding assay utilizes the radiolabeled soluble binding partner, such as a soluble binding partner/Fc fusion protein, and intact cells expressing the IMX polypeptide. Qualitative results can be obtained by competitive autoradiographic plate binding assays, while Scatchard plots (Scatchard, *Ann. N.Y. Acad. Sci.* 51:660, 1949) may be utilized to generate quantitative results.

## USE OF IMX NUCLEIC ACIDS OR OLIGONUCLEOTIDES

In addition to being used to express polypeptides as described above, the nucleic acids of the invention, including DNA, and oligonucleotides thereof can be used:

- as probes to identify nucleic acid encoding proteins homologous to IMX polypeptides;
- to identify human chromosomes;
- to map genes on human chromosome numbers 7, 19, and 22;
- to identify genes associated with certain diseases, syndromes, or other conditions associated with human chromosome numbers 7, 19, and 22;
- as single-stranded sense or antisense oligonucleotides, to inhibit expression of polypeptide encoded by the IMX sequences;
- to help detect defective genes in an individual; and
- for gene therapy.

Probes

Among the uses of nucleic acids of the invention is the use of fragments as probes or primers. Such fragments generally comprise at least about 17 contiguous nucleotides of a DNA sequence. In other embodiments, a DNA fragment comprises at least 30, or at least 60, contiguous nucleotides of a DNA sequence.

Because homologs of SEQ ID NOs:1-26, from other mammalian species, are contemplated herein, probes based on the human DNA sequence of SEQ ID NOs:1-26 may be used to screen cDNA libraries derived from other mammalian species, using conventional cross-species hybridization techniques.

Using knowledge of the genetic code in combination with the amino acid sequences set forth above, sets of degenerate oligonucleotides can be prepared. Such oligonucleotides are useful as primers, e.g., in polymerase chain reactions (PCR), whereby DNA fragments are isolated and amplified.

Identifying Chromosome Number

All or a portion of the nucleic acids of SEQ ID NOs:1-26, including oligonucleotides, can be used by those skilled in the art using well-known techniques to identify the human chromosomes, and the specific locus thereof, that contain the DNA of IMX family members. Useful techniques include, but are not limited to, using the sequence or portions, including oligonucleotides, as a probe in various well-known techniques such as in situ hybridization to chromosome spreads, Southern blot hybridization to hybrid cell lines, fluorescent tagging, and radiation hybrid mapping.

For example, chromosomes can be mapped by radiation hybridization. PCR is performed using the Whitehead Institute/MIT Center for Genome Research Genebridge4 panel of 93 radiation hybrids (http://www-genome.wi.mit.edu/ftp/distribution/human_STS_releases/july97/rhmap/genebridge4.html). Primers are used which lie within a putative exon of the gene of interest and which amplify a product from human genomic DNA, but do not amplify hamster genomic DNA. The results of the PCRs are converted into a data vector that is submitted to the Whitehead/MIT Radiation Mapping site on the internet (http://www-seq.wi.mit.edu). The data is scored and the chromosomal assignment and placement relative to known Sequence Tag Site (STS) markers on the radiation hybrid map is provided. The following web site provides additional information about radiation hybrid mapping: http://www-genome.wi.mit.edu/ftp/distribution/human_STS_releases/july97/07-97.INTRO.html).

Identifying Associated Diseases

As described previously, IMX molecules numbered 4, 21, 44, and 56 have been mapped to particular chromosome locations. Thus, the nucleic acid of a particular IMX molecule or a fragment thereof can be used by one skilled in the art using well-known techniques to analyze abnormalities associated with gene mapping to such chromosomes. This enables one to distinguish conditions in which this marker is rearranged or deleted. In addition, nucleotides of such IMX molecules or fragments thereof can be used as a positional marker to map other genes of previously unknown location.

The DNA may be used in developing treatments for any disorder mediated (directly or indirectly) by defective, or insufficient amounts of, the genes corresponding to the nucleic acids of the invention. Disclosure herein of native nucleotide sequences permits the detection of defective genes, and the replacement thereof with normal genes. Defective genes may be

detected in *in vitro* diagnostic assays, and by comparison of a native nucleotide sequence disclosed herein with that of a gene derived from a person suspected of harboring a defect in this gene.

Sense-Antisense

Other useful fragments of the nucleic acids include antisense or sense oligonucleotides comprising a single-stranded nucleic acid sequence (either RNA or DNA) capable of binding to target mRNA (sense) or DNA (antisense) sequences. Antisense or sense oligonucleotides according to the present invention comprise a fragment of DNA (SEQ ID NOs:1-26). Such a fragment generally comprises at least about 14 nucleotides, preferably from about 14 to about 30 nucleotides. The ability to derive an antisense or a sense oligonucleotide, based upon a cDNA sequence encoding a given protein is described in, for example, Stein and Cohen (*Cancer Res.* 48:2659, 1988) and van der Krol et al. (*BioTechniques* 6:958, 1988).

Binding of antisense or sense oligonucleotides to target nucleic acid sequences results in the formation of duplexes that block or inhibit protein expression by one of several means, including enhanced degradation of the mRNA by RNAseH, inhibition of splicing, premature termination of transcription or translation, or by other means. The antisense oligonucleotides thus may be used to block expression of proteins. Antisense or sense oligonucleotides further comprise oligonucleotides having modified sugar-phosphodiester backbones (or other sugar linkages, such as those described in WO91/06629) and wherein such sugar linkages are resistant to endogenous nucleases. Such oligonucleotides with resistant sugar linkages are stable *in vivo* (*i.e.*, capable of resisting enzymatic degradation) but retain sequence specificity to be able to bind to target nucleotide sequences.

Other examples of sense or antisense oligonucleotides include those oligonucleotides which are covalently linked to organic moieties, such as those described in WO 90/10448, and other moieties that increases affinity of the oligonucleotide for a target nucleic acid sequence, such as poly-(L-lysine). Further still, intercalating agents, such as ellipticine, and alkylating agents or metal complexes may be attached to sense or antisense oligonucleotides to modify binding specificities of the antisense or sense oligonucleotide for the target nucleotide sequence.

Antisense or sense oligonucleotides may be introduced into a cell containing the target nucleic acid sequence by any gene transfer method, including, for example, lipofection, $CaPO_4$-mediated DNA transfection, electroporation, or by using gene transfer vectors such as Epstein-Barr virus.

Sense or antisense oligonucleotides also may be introduced into a cell containing the target nucleotide sequence by formation of a conjugate with a ligand binding molecule, as described in WO 91/04753. Suitable ligand binding molecules include, but are not limited to, cell surface receptors, growth factors, other cytokines, or other ligands that bind to cell surface receptors. Preferably, conjugation of the ligand binding molecule does not substantially interfere with the ability of the ligand binding molecule to bind to its corresponding molecule or receptor, or block entry of the sense or antisense oligonucleotide or its conjugated version into the cell.

Alternatively, a sense or an antisense oligonucleotide may be introduced into a cell containing the target nucleic acid sequence by formation of an oligonucleotide-lipid complex, as described in WO 90/10448. The sense or antisense oligonucleotide-lipid complex is preferably dissociated within the cell by an endogenous lipase.

## USE OF IMX POLYPEPTIDES AND FRAGMENTED POLYPEPTIDES

Uses include, but are not limited to, the following:

- Purifying proteins and measuring activity thereof
- Delivery Agents
- Therapeutic and Research Reagents
- Molecular weight and Isoelectric focusing markers
- Controls for peptide fragmentation
- Identification of unknown proteins
- Preparation of Antibodies

### Purification Reagents

Each of the polypeptides of the invention finds use as a protein purification reagent. The polypeptides may be attached to a solid support material and used to purify (binding partner) proteins by affinity chromatography. In particular embodiments, a polypeptide (in any form described herein that is capable of binding (binding partner)) is attached to a solid support by conventional procedures. As one example, chromatography columns containing functional groups that will react with functional groups on amino acid side chains of proteins are available (Pharmacia Biotech, Inc., Piscataway, NJ). In an alternative, a polypeptide/Fc protein (as discussed above) is attached to Protein A- or Protein G-containing chromatography columns through interaction with the Fc moiety.

The polypeptide also finds use in purifying or identifying cells that express (binding partner) on the cell surface. Polypeptides are bound to a solid phase such as a column chromatography matrix or a similar suitable substrate. For example, magnetic microspheres can be coated with the polypeptides and held in an incubation vessel through a magnetic field. Suspensions of cell mixtures containing (binding partner) expressing cells are contacted with the solid phase having the polypeptides thereon. Cells expressing (binding partner) on the cell surface bind to the fixed polypeptides, and unbound cells then are washed away.

Alternatively, the polypeptides can be conjugated to a detectable moiety, then incubated with cells to be tested for (binding partner) expression. After incubation, unbound labeled matter is removed and the presence or absence of the detectable moiety on the cells is determined.

In a further alternative, mixtures of cells suspected of containing (binding partner) cells are incubated with biotinylated polypeptides. Incubation periods are typically at least one hour in duration to ensure sufficient binding. The resulting mixture then is passed through a column packed with avidin-coated beads, whereby the high affinity of biotin for avidin provides binding of the desired cells to the beads. Procedures for using avidin-coated beads are known (see Berenson, et al. *J. Cell. Biochem.*, 10D:239, 1986). Washing to remove unbound material, and the release of the bound cells, are performed using conventional methods.

Measuring Activity

Polypeptides also find use in measuring the biological activity of (binding partner) protein in terms of their binding affinity. The polypeptides thus may be employed by those conducting "quality assurance" studies, e.g., to monitor shelf life and stability of protein under different conditions. For example, the polypeptides may be employed in a binding affinity study to measure the biological activity of a (binding partner) protein that has been stored at different temperatures, or produced in different cell types. The proteins also may be used to determine whether biological activity is retained after modification of a (binding partner) protein (e.g., chemical modification, truncation, mutation, etc.). The binding affinity of the modified (binding partner) protein is compared to that of an unmodified (binding partner) protein to detect any adverse impact of the modifications on biological activity of (binding partner). The biological activity of a (binding partner) protein thus can be ascertained before it is used in a research study, for example.

Delivery Agents

The polypeptides also find use as carriers for delivering agents attached thereto to cells bearing binding partner. Cells expressing (binding partner) include those identified in (add citation if reference known). The polypeptides thus can be used to deliver diagnostic or therapeutic agents to such cells (or to other cell types found to express (binding partner) on the cell surface) in *in vitro* or *in vivo* procedures.

Detectable (diagnostic) and therapeutic agents that may be attached to a polypeptide include, but are not limited to, toxins, other cytotoxic agents, drugs, radionuclides, chromophores, enzymes that catalyze a colorimetric or fluorometric reaction, and the like, with the particular agent being chosen according to the intended application. Among the toxins are ricin, abrin, diphtheria toxin, *Pseudomonas aeruginosa* exotoxin A, ribosomal inactivating proteins, mycotoxins such as trichothecenes, and derivatives and fragments (e.g., single chains) thereof. Radionuclides suitable for diagnostic use include, but are not limited to, $^{123}I$, $^{131}I$, $^{99m}Tc$, $^{111}In$, and $^{76}Br$. Examples of radionuclides suitable for therapeutic use are $^{131}I$, $^{211}At$, $^{77}Br$, $^{186}Re$, $^{188}Re$, $^{212}Pb$, $^{212}Bi$, $^{109}Pd$, $^{64}Cu$, and $^{67}Cu$.

Such agents may be attached to the polypeptide by any suitable conventional procedure. The polypeptide comprises functional groups on amino acid side chains that can be reacted with functional groups on a desired agent to form covalent bonds, for example. Alternatively, the protein or agent may be derivatized to generate or attach a desired reactive functional group. The derivatization may involve attachment of one of the bifunctional coupling reagents available for attaching various molecules to proteins (Pierce Chemical Company, Rockford, Illinois). A number of techniques for radiolabeling proteins are known. Radionuclide metals may be attached to polypeptides by using a suitable bifunctional chelating agent, for example.

Conjugates comprising polypeptides and a suitable diagnostic or therapeutic agent (preferably covalently linked) are thus prepared. The conjugates are administered or otherwise employed in an amount appropriate for the particular application.

Therapeutic Agents

Polypeptides of the invention may be used in developing treatments for any disorder mediated (directly or indirectly) by defective, or insufficient amounts of the polypeptides. These polypeptides may be administered to a mammal afflicted with such a disorder.

The polypeptides may also be employed in inhibiting a biological activity of (binding partner), in *in vitro* or *in vivo* procedures. For example, a purified polypeptide may be used to

inhibit binding of (binding partner) to endogenous cell surface (binding partner). Biological effects that result from the binding of (binding partner) to endogenous receptors thus are inhibited.

IMX polypeptides may be administered to a mammal to treat a (binding partner-mediated disorder. Such (binding partner)-mediated disorders include conditions caused (directly or indirectly) or exacerbated by (binding partner).

Compositions of the present invention may contain a polypeptide in any form described herein, such as native proteins, variants, derivatives, oligomers, and biologically active fragments. In particular embodiments, the composition comprises a soluble polypeptide or an oligomer comprising soluble polypeptides.

Compositions comprising an effective amount of a polypeptide of the present invention, in combination with other components such as a physiologically acceptable diluent, carrier, or excipient, are provided herein. The polypeptides can be formulated according to known methods used to prepare pharmaceutically useful compositions. They can be combined in admixture, either as the sole active material or with other known active materials suitable for a given indication, with pharmaceutically acceptable diluents (e.g., saline, Tris-HCl, acetate, and phosphate buffered solutions), preservatives (e.g., thimerosal, benzyl alcohol, parabens), emulsifiers, solubilizers, adjuvants and/or carriers. Suitable formulations for pharmaceutical compositions include those described in *Remington's Pharmaceutical Sciences*, 16th ed. 1980, Mack Publishing Company, Easton, PA.

In addition, such compositions can be complexed with polyethylene glycol (PEG), metal ions, or incorporated into polymeric compounds such as polyacetic acid, polyglycolic acid, hydrogels, dextran, etc., or incorporated into liposomes, microemulsions, micelles, unilamellar or multilamellar vesicles, erythrocyte ghosts or spheroblasts. Such compositions will influence the physical state, solubility, stability, rate of *in vivo* release, and rate of *in vivo* clearance, and are thus chosen according to the intended application.

The compositions of the invention can be administered in any suitable manner, e.g., topically, parenterally, or by inhalation. The term "parenteral" includes injection, e.g., by subcutaneous, intravenous, or intramuscular routes, also including localized administration, e.g., at a site of disease or injury. Sustained release from implants is also contemplated. One skilled in the pertinent art will recognize that suitable dosages will vary, depending upon such factors as the nature of the disorder to be treated, the patient's body weight, age, and general condition, and the route of administration. Preliminary doses can be determined according to animal tests, and

the scaling of dosages for human administration is performed according to art-accepted practices.

Compositions comprising nucleic acids in physiologically acceptable formulations are also contemplated. DNA may be formulated for injection, for example.

Research Agents

Another use of the polypeptide of the present invention is as a research tool for studying the biological effects that result from inhibiting interactions between polypeptides of the invention and their "binding partner". Polypeptides also may be employed in *in vitro* assays for detecting binding partner or cells expressing the binding partner or the interactions thereof.

Molecular Weight, Isoelectric Point Markers

The polypeptides of the present invention can be subjected to fragmentation into smaller peptides by chemical and enzymatic means, and the peptide fragments so produced can be used in the analysis of other proteins or polypeptides. For example, such peptide fragments can be used as peptide molecular weight markers, peptide isoelectric point markers, or in the analysis of the degree of peptide fragmentation. Thus, the invention also includes these polypeptides and peptide fragments, as well as kits to aid in the determination of the apparent molecular weight and isoelectric point of an unknown protein and kits to assess the degree of fragmentation of an unknown protein.

Although all methods of fragmentation are encompassed by the invention, chemical fragmentation is a preferred embodiment, and includes the use of cyanogen bromide to cleave under neutral or acidic conditions such that specific cleavage occurs at methionine residues (E. Gross, *Methods in Enz.* 11:238-255, 1967). This can further include additional steps, such as a carboxymethylation step to convert cysteine residues to an unreactive species.

Enzymatic fragmentation is another preferred embodiment, and includes the use of a protease such as Asparaginylendo-peptidase, Arginylendo-peptidase, *Achromobacter* protease I, Trypsin, *Staphlococcus aureus* V8 protease, Endoproteinase Asp-N, or Endoproteinase Lys-C under conventional conditions to result in cleavage at specific amino acid residues. Asparaginylendo-peptidase can cleave specifically on the carboxyl side of the asparagine residues present within the polypeptides of the invention. Arginylendo-peptidase can cleave specifically on the carboxyl side of the arginine residues present within these polypeptides. *Achromobacter* protease I can cleave specifically on the carboxyl side of the lysine residues present within the polypeptides (Sakiyama and Nakat, U.S. Patent No. 5,248,599; T. Masaki et al., *Biochim. Biophys. Acta 660*:44-50, 1981; T. Masaki et al., *Biochim. Biophys. Acta 660*:51-

55, 1981). Trypsin can cleave specifically on the carboxyl side of the arginine and lysine residues present within polypeptides of the invention. Enzymatic fragmentation may also occur with a protease that cleaves at multiple amino acid residues. For example, *Staphlococcus aureus* V8 protease can cleave specifically on the carboxyl side of the aspartic and glutamic acid residues present within polypeptides (D. W. Cleveland, *J. Biol. Chem.* 3:1102-1106, 1977). Endoproteinase Asp-N can cleave specifically on the amino side of the asparagine residues present within polypeptides. Endoproteinase Lys-C can cleave specifically on the carboxyl side of the lysine residues present within polypeptides of the invention. Other enzymatic and chemical treatments can likewise be used to specifically fragment these polypeptides into a unique set of specific peptides.

Of course, the peptides and fragments of the polypeptides of the invention can also be produced by conventional recombinant processes and synthetic processes well known in the art. With regard to recombinant processes, the polypeptides and peptide fragments encompassed by invention can have variable molecular weights, depending upon the host cell in which they are expressed. Glycosylation of polypeptides and peptide fragments of the invention in various cell types can result in variations of the molecular weight of these pieces, depending upon the extent of modification. The size of these pieces can be most heterogeneous with fragments of polypeptide derived from the extracellular portion of the polypeptide. Consistent polypeptides and peptide fragments can be obtained by using polypeptides derived entirely from the transmembrane and cytoplasmic regions, pretreating with N-glycanase to remove glycosylation, or expressing the polypeptides in bacterial hosts.

The molecular weight of these polypeptides can also be varied by fusing additional peptide sequences to both the amino and carboxyl terminal ends of polypeptides of the invention. Fusions of additional peptide sequences at the amino and carboxyl terminal ends of polypeptides of the invention can be used to enhance expression of these polypeptides or aid in the purification of the protein. In addition, fusions of additional peptide sequences at the amino and carboxyl terminal ends of polypeptides of the invention will alter some, but usually not all, of the fragmented peptides of the polypeptides generated by enzymatic or chemical treatment. Of course, mutations can be introduced into polypeptides of the invention using routine and known techniques of molecular biology. For example, a mutation can be designed so as to eliminate a site of proteolytic cleavage by a specific enzyme or a site of cleavage by a specific chemically induced fragmentation procedure. The elimination of the site will alter the peptide fingerprint of polypeptides of the invention upon fragmentation with the specific enzyme or chemical procedure.

The polypeptides and the resultant fragmented peptides can be analyzed by methods including sedimentation, electrophoresis, chromatography, and mass spectrometry to determine their molecular weights. Because the unique amino acid sequence of each piece specifies a molecular weight, these pieces can thereafter serve as molecular weight markers using such analysis techniques to assist in the determination of the molecular weight of an unknown protein, polypeptides or fragments thereof. The molecular weight markers of the invention serve particularly well as molecular weight markers for the estimation of the apparent molecular weight of proteins that have similar apparent molecular weights and, consequently, allow increased accuracy in the determination of apparent molecular weight of proteins.

When the invention relates to the use of fragmented peptide molecular weight markers, those markers are preferably at least 10 amino acids in size. More preferably, these fragmented peptide molecular weight markers are between 10 and 100 amino acids in size. Even more preferable are fragmented peptide molecular weight markers between 10 and 50 amino acids in size and especially between 10 and 35 amino acids in size. Most preferable are fragmented peptide molecular weight markers between 10 and 20 amino acids in size.

Among the methods for determining molecular weight are sedimentation, gel electrophoresis, chromatography, and mass spectrometry. A particularly preferred embodiment is denaturing polyacrylamide gel electrophoresis (U. K. Laemmli, *Nature 227*:680-685, 1970). Conventionally, the method uses two separate lanes of a gel containing sodium dodecyl sulfate and a concentration of acrylamide between 6-20%. The ability to simultaneously resolve the marker and the sample under identical conditions allows for increased accuracy. It is understood, of course, that many different techniques can be used for the determination of the molecular weight of an unknown protein using polypeptides of the invention, and that this embodiment in no way limits the scope of the invention.

Each unglycosylated polypeptide or fragment thereof has a pI that is intrinsically determined by its unique amino acid sequence (which pI can be estimated by the skilled artisan using any of the computer programs designed to predict pI values currently available, calculated using any well-known amino acid pKa table, or measured empirically). Therefore these polypeptides and fragments thereof can serve as specific markers to assist in the determination of the isoelectric point of an unknown protein, polypeptide, or fragmented peptide using techniques such as isoelectric focusing. These polypeptide or fragmented peptide markers serve particularly well for the estimation of apparent isoelectric points of unknown proteins that have apparent isoelectric points close to that of the polypeptide or fragmented peptide markers of the invention.

The technique of isoelectric focusing can be further combined with other techniques such as gel electrophoresis to simultaneously separate a protein on the basis of molecular weight and charge. The ability to simultaneously resolve these polypeptide or fragmented peptide markers and the unknown protein under identical conditions allows for increased accuracy in the determination of the apparent isoelectric point of the unknown protein. This is of particular interest in techniques, such as two dimensional electrophoresis (T.D. Brock and M.T. Madigan, *Biology of Microorganisms* 76-77 (Prentice Hall, 6d ed. 1991)), where the nature of the procedure dictates that any markers should be resolved simultaneously with the unknown protein. In addition, with such methods, these polypeptides and fragmented peptides thereof can assist in the determination of both the isoelectric point and molecular weight of an unknown protein or fragmented peptide.

Polypeptides and fragmented peptides can be visualized using two different methods that allow a discrimination between the unknown protein and the molecular weight markers. In one embodiment, the polypeptide and fragmented peptide molecular weight markers of the invention can be visualized using antibodies generated against these markers and conventional immunoblotting techniques. This detection is performed under conventional conditions that do not result in the detection of the unknown protein. It is understood that it may not be possible to generate antibodies against all polypeptide fragments of the invention, since small peptides may not contain immunogenic epitopes. It is further understood that not all antibodies will work in this assay; however, those antibodies which are able to bind polypeptides and fragments of the invention can be readily determined using conventional techniques.

The unknown protein is also visualized by using a conventional staining procedure. The molar excess of unknown protein to polypeptide or fragmented peptide molecular weight markers of the invention is such that the conventional staining procedure predominantly detects the unknown protein. The level of these polypeptide or fragmented peptide molecular weight markers is such as to allow little or no detection of these markers by the conventional staining method. The preferred molar excess of unknown protein to polypeptide molecular weight markers of the invention is between 2 and 100,000 fold. More preferably, the preferred molar excess of unknown protein to these polypeptide molecular weight markers is between 10 and 10,000 fold and especially between 100 and 1,000 fold.

It is understood of course that many techniques can be used for the determination and detection of molecular weight and isoelectric point of an unknown protein, polypeptides, and fragmented peptides thereof using these polypeptide molecular weight markers and peptide fragments thereof and that these embodiments in no way limit the scope of the invention.

In another embodiment, the analysis of the progressive fragmentation of the polypeptides of the invention into specific peptides (D. W. Cleveland et al., *J. Biol. Chem.* 252:1102-1106, 1977), such as by altering the time or temperature of the fragmentation reaction, can be used as a control for the extent of cleavage of an unknown protein. For example, cleavage of the same amount of polypeptide and unknown protein under identical conditions can allow for a direct comparison of the extent of fragmentation. Conditions that result in the complete fragmentation of the polypeptide can also result in complete fragmentation of the unknown protein.

As to the specific use of the polypeptides and fragmented peptides of the invention as molecular weight markers, the fragmentation of the polypeptides of SEQ ID NOs:27-38 with cyanogen bromide generates a unique set of fragmented peptide molecular weight markers. An additional fragment results if the initiating methionine is present. The distribution of methionine residues determines the number of amino acids in each peptide and the unique amino acid composition of each peptide determines its molecular weight.

In addition, the preferred purified polypeptides of the invention (SEQ ID NOs:27-38) have calculated molecular weights of approximately 3683, 1783, 11248, 75503, 43040, 8051, 33306, 3515, 10736, 25162, and 2450 Daltons.

Where an intact protein is used, the use of these polypeptide molecular weight markers allows increased accuracy in the determination of apparent molecular weight of proteins that have apparent molecular weights close to 3683, 1783, 11248, 75503, 43040, 8051, 33306, 3515, 10736, 25162, or 2450 Daltons. Where fragments are used, there is increased accuracy in determining molecular weight over the range of the molecular weights of the fragment.

Finally, as to the kits that are encompassed by the invention, the constituents of such kits can be varied, but typically contain the polypeptide and fragmented peptide molecular weight markers. Also, such kits can contain the polypeptides wherein a site necessary for fragmentation has been removed. Furthermore, the kits can contain reagents for the specific cleavage of the polypeptide and the unknown protein by chemical or enzymatic cleavage. Kits can further contain antibodies directed against polypeptides or fragments thereof of the invention.

Identification of Unknown Proteins

As set forth above, a polypeptide or peptide fingerprint can be entered into or compared to a database of known proteins to assist in the identification of the unknown protein using mass spectrometry (W.J. Henzel et al., *Proc. Natl. Acad. Sci. USA* 90:5011-5015, 1993; D. Fenyo et al., *Electrophoresis* 19:998-1005, 1998). A variety of computer software programs to facilitate these comparisons are accessible via the Internet, such as Protein Prospector (Internet site:

prospector.uscf.edu). MultiIdent (Internet site: www.expasy.ch/sprot/multiident.html), PeptideSearch (Internet site: www.mann.embl-heiedelberg.de...deSearch/FR_PeptideSearch Form.html), and ProFound (Internet site: www.chait-sgi.rockefeller.edu/cgi-bin/prot-id-frag.html). These programs allow the user to specify the cleavage agent and the molecular weights of the fragmented peptides within a designated tolerance. The programs compare these molecular weights to protein databases to assist in determining the identity of the unknown protein.

In addition, a polypeptide or peptide digest can be sequenced using tandem mass spectrometry (MS/MS) and the resulting sequence searched against databases (J.K. Eng, et al., *J. Am. Soc. Mass Spec.* 5:976-989 (1994); M. Mann and M. Wilm, *Anal. Chem.* 66:4390-4399 (1994); J.A. Taylor and R.S. Johnson, *Rapid Comm. Mass Spec.* 11:1067-1075 (1997)). Searching programs that can be used in this process exist on the Internet, such as Lutefisk 97 (Internet site: www.lsbc.com:70/Lutefisk97.html), and the Protein Prospector, Peptide Search and ProFound programs described above. Therefore, adding the sequence of a gene and its predicted protein sequence and peptide fragments to a sequence database can aid in the identification of unknown proteins using tandem mass spectrometry.

Antibodies

Antibodies that are immunoreactive with the polypeptides of the invention are provided herein. Such antibodies specifically bind to the polypeptides *via* the antigen-binding sites of the antibody (as opposed to non-specific binding). Thus, the polypeptides, fragments, variants, fusion proteins, etc., as set forth above may be employed as immunogens in producing antibodies immunoreactive therewith.

Polyclonal and monoclonal antibodies may be prepared by conventional techniques. See, for example, *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Kennet et al. (eds.), Plenum Press, New York (1980); and *Antibodies: A Laboratory Manual* , Harlow and Land (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1988).

Antigen-binding fragments of such antibodies, which may be produced by conventional techniques, are also encompassed by the present invention. Examples of such fragments include, but are not limited to, Fab and $F(ab')_2$ fragments. Antibody fragments and derivatives produced by genetic engineering techniques are also provided.

The monoclonal antibodies of the present invention include chimeric antibodies, e.g., humanized versions of murine monoclonal antibodies. Such humanized antibodies may be

prepared by known techniques, and offer the advantage of reduced immunogenicity when the antibodies are administered to humans. In one embodiment, a humanized monoclonal antibody comprises the variable region of a murine antibody (or just the antigen binding site thereof) and a constant region derived from a human antibody. Alternatively, a humanized antibody fragment may comprise the antigen binding site of a murine monoclonal antibody and a variable region fragment (lacking the antigen-binding site) derived from a human antibody. Procedures for the production of chimeric and further engineered monoclonal antibodies include those described in Riechmann et al. (*Nature 332*:323, 1988), Liu et al. (*PNAS 84*:3439, 1987), Larrick et al. (*Bio/Technology 7*:934, 1989), and Winter and Harris (*TIPS 14*:139, May, 1993). Procedures to generate antibodies transgenically can be found in GB 2,272,440, US Patent Nos. 5,569,825 and 5,545,806 and related patents claiming priority therefrom, all of which are incorporated by reference herein.

In one embodiment, the antibodies are specific for the polypeptides of the present invention, and do not cross-react with other proteins. Screening procedures by which such antibodies may be identified are well known, and may involve immunoaffinity chromatography, for example.

Hybridoma cell lines that produce monoclonal antibodies specific for the polypeptides of the invention are also contemplated herein. Such hybridomas may be produced and identified by conventional techniques. One method for producing such a hybridoma cell line comprises immunizing an animal with a polypeptide; harvesting spleen cells from the immunized animal; fusing said spleen cells to a myeloma cell line, thereby generating hybridoma cells; and identifying a hybridoma cell line that produces a monoclonal antibody that binds the polypeptide. The monoclonal antibodies may be recovered by conventional techniques.

Uses Thereof

The antibodies of the invention can be used in assays to detect the presence of the polypeptides or fragments of the invention, either *in vitro* or *in vivo*. The antibodies also may be employed in purifying polypeptides or fragments of the invention by immunoaffinity chromatography.

Those antibodies that additionally can block binding of the polypeptides of the invention to the binding partner may be used to inhibit a biological activity that results from such binding. Such blocking antibodies may be identified using any suitable assay procedure, such as by testing antibodies for the ability to inhibit binding of the binding partner to certain cells expressing the binding partner. Alternatively, blocking antibodies may be identified in assays

for the ability to inhibit a biological effect that results from binding of the binding partner to target cells.

Such an antibody may be employed in an *in vitro* procedure, or administered *in vivo* to inhibit a biological activity mediated by the entity that generated the antibody. Disorders caused or exacerbated (directly or indirectly) by the interaction of (binding partner) with cell surface (binding partner) receptor thus may be treated. A therapeutic method involves *in vivo* administration of a blocking antibody to a mammal in an amount effective in inhibiting a (binding partner)-mediated biological activity. Monoclonal antibodies are generally preferred for use in such therapeutic methods. In one embodiment, an antigen-binding antibody fragment is employed.

Antibodies may be screened for agonistic (*i.e.*, ligand-mimicking) properties. Such antibodies, upon binding to cell surface antigen, induce biological effects (e.g., transduction of biological signals) similar to the biological effects induced when (binding partner) binds to cell surface antigen.

Compositions comprising an antibody that is directed against a polypeptide of the invention, and a physiologically acceptable diluent, excipient, or carrier, are provided herein.

Also provided herein are conjugates comprising a detectable (e.g., diagnostic) or therapeutic agent, attached to the antibody. Examples of such agents are presented above. The conjugates find use in *in vitro* or *in vivo* procedures.

The following examples are provided to further illustrate particular embodiments of the invention, and are not to be construed as limiting the scope of the present invention.

## EXAMPLE 1: Isolation of the IMX Nucleic Acids

### The T84 Epithelial Barrier Model

As discussed above, damage to the intestinal epithelial barrier is a hallmark of (IBD), and a number of *in vitro* models of epithelial barrier function have been developed over the years. The best characterized of these models is the T84 intestinal epithelial barrier system, Dharmsathaphorn et al., *Am. J. Physiol.*, 246:G204-G208, 1984 and Madara et al., *J. Cell Biol.*, 101:2124-2133,1985).

T84 cells were plated on 75 mm polycarbonate transwell filter inserts (Costar) and grown in DME/F12 (1:1) containing 10% heat-inactivated bovine calf serum. The cells were maintained at confluence for 2-3 days, and integrity of the epithelial barrier was determined by measuring transepithelial electrical resistance (TER) using an EVOM epithelial voltohmmeter (World Precision Instruments). When the TER values were greater than 1000 ohms/cm2 and

were stable, cells were treated with interferon-g (30 ng/ml. Genzyme) added to the basolateral side of the membrane. At various times after treatment (4, 24 and 44h), TERs were measured to monitor the interferon-induced disruption of the barrier, and RNA was harvested from the cells at those time points using TRIzol reagent (Life Technologies). RNA was extracted using conventional methods and subsequently used for TOGA™ analysis as described in Example 2.

TOGA™ analysis

This example describes a method for determining mRNA expression characteristics. The isolated RNA was analyzed using a method of simultaneous sequence-specific identification of mRNAs known as TOGA™ (Total Gene expression Analysis) described in U.S. Patent No. 5,459,037 and U.S. Patent No. 5,807,680, hereby incorporated herein by reference. Preferably, prior to the application of the TOGA™ technique, the isolated RNA was enriched to form a starting polyA-containing mRNA population by methods known in the art. In a preferred embodiment, the TOGA™ method further comprised an additional PCR step performed using four separate reactions, one for each of the four 5' PCR primers, and cDNA templates prepared from a population of antisense cRNAs. A final PCR step used 256 5' PCR primers in 64 subpools for each of the four reactions of the previous step produced PCR products that were cDNA fragments that corresponded to the 3'-region of the starting mRNA population.

The produced PCR products were then identified by a) the sequence of at least the 5' seven base pairs, preferably the sequence of the entire fragment, and b) the length of the fragment. These two parameters, sequence and fragment length, were used to compare the obtained PCR products to a database of known polynucleotide sequences. A database search for homologous sequences in Genbank resulted in no matches, indicating the novelty of the IMX sequences of the invention. The intensities of the PCR products were compared across the 4 input RNA samples (t=0, 4hr, 24hr, 44hr) and species that were regulated were identified and further characterized (Table 1).

The DSTs of SEQ ID NO:1-26 and fragments thereof are useful as probes to study and diagnose the changes in gene expression demonstrated by the data of Table 1. Alternatively, polypeptides and fragments thereof that are the translations of SEQ ID NO:1-26 and fragments thereof are useful as probes to study and diagnose the changes in gene expression demonstrated by the data of Table 1.

| | | | Relative PCR Fragment Amount | | | |
|---|---|---|---|---|---|---|
| Name | DST SEQ ID NO: | Digital Address (Msp I) | 0 | 4hr | 24hr | 44hr |
| IMX 4 | 1 | TAAG 187 | 41 | 94 | 350 | 299 |
| IMX 10 | 2 | TTCT 159 | 250 | 716 | 250 | 207 |
| IMX 21 | 3 | CTTC 364 | 35 | 60 | 330 | 749 |
| IMX 28 | 4 | CCAT 414 | 138 | 435 | 647 | 464 |
| IMX 32 | 5 | CCGA 374 | 237 | 539 | 25 | 32 |
| IMX 39 | 6 | TGGA 450 | 24 | 87 | 110 | 78 |
| IMX 40 | 7 | TCTA 163 | 38 | 294 | 67 | 153 |
| IMX 42 | 8 | GAGC 426 | 145 | 733 | 876 | 1076 |
| IMX 44 | 9 | CTGC 221 | 248 | 344 | 1499 | 1624 |
| IMX 56 | 10 | GATA 111 | 955 | 1223 | 2357 | 3450 |

Table 1. IMX Nucleic Acid Molecules. Relative expression levels determined by TOGA™.

## EXAMPLE 2: Further Characterization of IMX4

### Nucleic Acid Molecules and Polypeptides

An additional 600bp sequence (SEQ ID NO:11) that included DST IMX4 (SEQ ID NO:1) was obtained by anchor PCR from a T84 library. A clone approximately 2.0kb in length was then isolated from a commercial library (Origene). The sequence of the 3' end of the clone (SEQ ID NO:12) included the sequence of the IMX4 DST. The sequence of the 5' end of the clone (SEQ ID NO:13) matched part of the human ApoL gene (AF019225, Figure 21). Complete sequence of the clone was not obtained. However, comparison of 5' end of the clone suggests it represents an alternative splice product to the reported ApoL sequence, i.e., bases 1-168 match to 2 exons on PAC carrying the ApoL gene but are not included in the reported complete cDNA. See Figure 22. A polypeptide translated from IMX4 is provided in SEQ ID NO: 27.

## EXAMPLE 3: Further Characterization of IMX10 and IMX21

### Nucleic Acid Molecules and Polypeptides

Further experiments failed to show regulation of these DSTs in replications using PCR with extended primers and Northern blots. A polypeptide translated from IMX10 (SEQ ID

NO:14) is provided in SEQ ID NO: 28. A polypeptide translated from IMX21 (SEQ ID NO:15) is provided in SEQ ID NO: 29.

## EXAMPLE 4: Further Characterization of IMX28

### Nucleic Acid Molecules and Polypeptides

One polypeptide translated from IMX28 is provided in SEQ ID NO: 30.

A GST fusion protein of IMX28 was used in studies designed to identify potential ligands for this enzyme.The GST fusion of IMX28 was prepared as follows: IMX28-8, a clone with the full length coding sequence, was excised from pGEM by SalI-NotI digestion and subcloned into SalI-NotI digested pGEX-5X-3 (Pharmacia). The pGEX-5X-3 plasmid contains the Schistosoma japonicum glutathione-S-transferase (GST) gene 5' to the SalI-NotI cloning site. Insertion of the IMX28-8 DNA fragment into this cloning site results in a fusion gene containing GST-IMX28 in the proper reading frame. The complete sequence of the fusion protein consists of GST followed by a Factor Xa tripeptide cleavage site, the peptide consisting of Gly-Ile-Pro-Arg-Asn-Ser-Arg-Val-Asp-Ala-Thr that is derived partially from the pGEX vector and partially from the SalI-NotI fragment of IMX28 and IMX28. The vector containing the fusion gene was electroporated into competent bacterial cells and a single colony containing the DNA, confirmed by diagnostic restriction digestion, was expanded to a 250ml culture for purification of the fusion protein.

Briefly, cells from these cultures were pelleted and resupended in STE buffer containing a protease inhibitor cocktail and 20ug/ml hen egg lysozyme. The cells were lysed by addition of 10% sarcosyl in STE (final sarcosyl concentration was 1.5%), sonicated and centrifuged to remove cell debris. The supernatant was passed over a Glutathione-Sepharose 4B column, washed sequentially with 1.5% sarcosyl in STE and then STE. The bound material, containing the GST-IMX28 fusion protein was eluted by adding 20mM Glutathione in STE. The eluted material was analyzed by SDS-PAGE to confirm the presence of the GST-IMX28 fusion protein.

A recombinant adenovirus construct expressing IMX28 was used to infect T84 cells to determine effects on epithelial barrier formation. To construct adenoviral vectors with untagged full length Imx28 and NH2-terminally FLAG tagged Imx28, the pADEASY1 system described by He et al. (1998 ) Proc. Natl. Acad. Sci. USA 95: 2509-2514, was used. For the untagged construct a Not1 fragment carrying the full coding region of Imx28 was inserted into the polylinker of the pAdtrackCMV shuttle vector. After confirmation of the correct orientation the

vector was cut with the restriction enzyme Pac1 and co-transformed into E. coli. Strain BJ5183 along with the pADEasy1 vector. Recombinants containing the Imx28 product were screened for as described and by gene specific PCR. To generate non-replicative, infectious viral particles the helper strain 293EBNA-EBNA or 293MSR was transfected with PmeI linearized recombinant vector. After 3 days virus was harvested by freeze-thaw lysis of the transfected cells. Several rounds of infection of these cells with viral supernatents were performed to boost titers. The tagged version was prepared by addition of the FLAG sequence and a Gly-Gly-Gly-Gly spacer region onto the full length coding sequence of Imx28 via PCR with a 5' oligo with the incorporated nucleotides for the tag and spacer. The PCR product was cloned into a TA- vector and excised with NotI and cloned as above. Screening and viral particle production was as above. No deleterious or positive effects were observed on T84 barrier function in the presence or absence of IFN-gamma.

## EXAMPLE 5: Further Characterization of IMX32

### Nucleic Acid Molecules and Polypeptides

The original clone produced by extending the IMX32 DST was 1588 bp in length (SEQ ID NO: 17). The first round of anchor PCR yielded a 670bp product that was used to produce a contig that extended the sequence 585 bases to 2173 (SEQ ID NO: 17). Another round of anchor PCR yielded the 1676 bp product, immunex32–a24.seq. (SEQ ID NO: 18) which further extended the contig for Imx32 to 2834 bp. In addition, the immunex32-a24.seq product revealed an alternative splice lacking bases 211-639 of IMX32-1 clone.

There is an ORF for 155 amino acids at the 5' end of the sequence (SEQ ID NO:31 translated from SEQ ID NO: 16; SEQ ID NO:32 translated from SEQ ID NO: 18). BLAST analysis suggests that the ORF may contain a KRAB-like domain because it has partial matches to Zn-finger proteins which contain KRAB domains at their N-terminus such as ZNF140.

## EXAMPLE 6: Further Characterization of IMX39

### Nucleic Acid Molecules and Polypeptides

One polypeptide translated from IMX39 is provided in SEQ ID NO: 33.

IMX39 FLAG-tagged and untagged adenoviral vector versions were prepared as outlined above for Imx28 in Example 4, above. Controls for the expression were infection with other FLAG- tagged adenoviral delivered proteins such as IMX5. Based upon the predicted structure from its cDNA sequence, it was expected that the IMX39 polypeptide would be a cytoplasmic protein. However, since some secreted cytokines, such as those of the IL-1 family, lack a

predicted signal sequence but are secreted from the cell, this expectation was tested experimentally in both T84 cells via adenoviral mediated transduction and via transfection into CV1EBNA cells with another expression vector, pDC412. Despite evidence for succesful expression in the transfected cells, no product was found in the media as determined by FLAG western (T84 system) or 35-S radiolabeled product (CV1EBNA system). The positive control in the CV1 system was an Imx44-Fc fusion protein known to be secrete. In both systems empty vectors were used as a negative control. The attempted expression of IMX39 polypeptides had no apparent effect on barrier function in the absence or presence of IFN-gamma.

## EXAMPLE 7: Further Characterization of IMX40

### Nucleic Acid Molecules and Polypeptides

Anchor PCR using an Origene human small intestine library revealed a product of 1669 bp (SEQ ID NO: 20). Bases 1-265 of the DST IMX40 (SEQ ID NO:7) align with bases 1565-1669 of this sequence. The anchor PCR product has a potential 155 aa ORF from bases 185-650. A translation of this ORF is given in SEQ ID NO:34. This ORF has no recognizable features, motifs, or database homologies.

## EXAMPLE 8: Further Characterization of IMX42

### Nucleic Acid Molecules and Polypeptides

Several attempts to screen libraries using DST IMX42 (SEQ ID NO:8) as well as anchor PCR and RACE have not produced a clone with a longer sequence. BLAST comparison against against imxhutdb extended sequence approx 200 bases to 592 bases (SEQ ID NO: 21). However, the sequence extension did not doesn't increase any database matches found by comparison to the DST IMX42 sequence. The contig has a potential ORF of 134 amino acids (SEQ ID NO: 36) which, while a COOH terminal extension of SEQ ID NO:35, may be a partial sequence.

## EXAMPLE 9: Further Characterization of IMX44

### Nucleic Acid Molecules and Polypeptides

Longer polynucleotides corresponding to DST IMX44 have been identified (SEQ ID NO:22, 23, 24). A translated polypeptide is found in SEQ ID NO:37. A soluble Fc form of IMX44 polypeptide was synthesized and used in various assays. IMX44-Fc had no effect on T84 barrier function in the absence or presence of IFN-gamma. IMX44-Fc had no effect on natural killer (NK) cell activation. IMX44-Fc had no positive hits on cognate screen assays.

A soluble FLAG polyHis form was also produced. No activity in cellular activation assays was found nor any alteration of cytokine production using this polypeptide in assays.

The expression of IMX44 in various murine models of gut inflammation was determined by Northern and array analysis. Little to no regulation of transcript was found was found in anti-CD3-induced ileitis in C57BL/6 mice, DSS-induced colitis in BALB/c mice or C57BL/6- mice, *mdr*1 knock out mice with colitis, and IFN-gamma stimulated LN T cells.

## EXAMPLE 10: Further Characterization of IMX56

### Nucleic Acid Molecules and Polypeptides

A translated IMX56 polypeptide is found in SEQ ID NO:38. Comparison of the DST sequence IMX56 to IMAGE consortium clones extended the sequence (SEQ ID NO:25) which was 3' on sequenced PAC to the described end of human ApoL. Anchor PCR using a T84 library produced results that indicated that the IMX56 DST is derived from an alternate 3' UTR of ApoL.

## EXAMPLE 11: Monoclonal Antibodies That Bind IMX Polypeptides

Monoclonal antibodies that bind the polypeptides of the invention can be prepared by methods well known in the art. Suitable immunogens that may be employed in generating such antibodies include, but are not limited to, purified IMX polypeptides or an immunogenic fragment thereof such as the extracellular domain, or fusion proteins containing IMX polypeptides (e.g., a soluble IMX 21polypeptide/Fc fusion protein).

Purified IMX polypeptides of the invention can be used to generate monoclonal antibodies immunoreactive therewith, using conventional techniques such as those described in U.S. Patent 4,411,993. Briefly, mice are immunized with an IMX polypeptide immunogen emulsified in complete Freund's adjuvant, and injected in amounts ranging from 10-100 :g

subcutaneously or intraperitoneally. Ten to twelve days later, the immunized animals are boosted with additional IMX polypeptide immunogen emulsified in incomplete Freund's adjuvant. Mice are periodically boosted thereafter on a weekly to bi-weekly immunization schedule. Serum samples are periodically taken by retro-orbital bleeding or tail-tip excision to test for anti-IMX polypeptide antibodies by dot blot assay, ELISA (Enzyme-Linked Immunosorbent Assay) or inhibition of IMX polypeptide/binding partner interactions.

Following detection of an appropriate antibody titer, positive animals are provided one last intravenous injection of IMX polypeptide immunogen in saline. Three to four days later, the animals are sacrificed, spleen cells harvested, and spleen cells are fused to a murine myeloma cell line, e.g., NS1 or preferably P3x63Ag8.653 (ATCC CRL 1580). Fusions generate hybridoma cells, which are plated in multiple microtiter plates in a HAT (hypoxanthine, aminopterin and thymidine) selective medium to inhibit proliferation of non-fused cells, myeloma hybrids, and spleen cell hybrids.

The hybridoma cells are screened by ELISA for reactivity against the purified IMX polypeptide of interest by adaptations of the techniques disclosed in Engvall et al., *Immunochem*. 8:871, 1971 and in U.S. Patent 4,703,004. A preferred screening technique is the antibody capture technique described in Beckmann et al., (*J. Immunol*. 144:4212, 1990) Positive hybridoma cells can be injected intraperitoneally into syngeneic BALB/c mice to produce ascites containing high concentrations of anti-IMX polypeptide monoclonal antibodies. Alternatively, hybridoma cells can be grown *in vitro* in flasks or roller bottles by various techniques. Monoclonal antibodies produced in mouse ascites can be purified by ammonium sulfate precipitation, followed by gel exclusion chromatography. Alternatively, affinity chromatography based upon binding of antibody to Protein A or Protein G can also be used, as can affinity chromatography based upon binding to the IMX polypeptide of interest.

Alternatively, IMX nucleic acid molecules or fragments thereof can be expressed to produce IMX polypeptides or fragments thereof, that can be used to make antibodies that are useful for identifying corresponding polypeptides in techniques such as western blotting, immunocytochemistry ,and ELISA assays using standard techniques such as those described in U.S. Patent No. 4,900,811, incorporated by reference herein.

The references cited herein are incorporated by reference herein in their entirety.